Model Theory

no. 2 vol. 3 2024

$M \models T$

Special issue on the occasion of the 75th birthday of

Boris Zilber

msp

Model Theory

msp.org/mt

See inside back cover or msp.org/mt for submission instructions.

Model Theory (ISSN 2832-904X electronic, 2832-9058 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840 is published continuously online.

MT peer review and production are managed by EditFlow® from MSP.

PUBLISHED BY
mathematical sciences publishers
nonprofit scientific publishing
https://msp.org/

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.199

Introduction

This volume comprises a mix of new research, retrospective reviews, and informal histories, all concerning or inspired by the contributions to model theory of our friend and ex-supervisor Boris Zilber, and dedicated to him in celebration of his 75th birthday. The varied subjects of the articles reflect aspects of Boris's mathematical focus at different stages of his career, which we can conveniently group around a number of influential conjectures. Boris's long-sighted vision is made nowhere clearer than in his formulation of conjectures which summarise our ignorance and point a way out of it. Each has inspired massive efforts, but none can yet be considered fully resolved.

A persistent theme of Boris's work has been the study of categorical structures. In the 1970s he developed a fine structure theory for uncountably and totally categorical structures. Along with such key results as the weak trichotomy theorem and the ladder theorem, this resulted in a pair of conjectures which are key to the first two themes of this volume. The *trichotomy conjecture* attempted a classification of the combinatorial geometries of strongly minimal sets, requiring a field structure to be responsible for any failure of (local) modularity. Although false, this conjecture remains fruitful. The articles of Castle and Hasson and of Chatzidakis and Hrushovski in this volume are directly concerned with situations in which the conjecture holds, and Moosa's article introduces a new theory in which it is strongly expected to hold. Kennedy's article is a philosophical take on the inspiration around the conjecture and Zilber's mathematics more generally. Related aspects of combinatorial geometries are discussed in Cameron's article. In recent decades, model theorists have become increasingly comfortable with extending ideas gestated in the categorical or ω-stable context to wider classes of theories, and this is on show in the articles of Ammer and Tent on strictly stable theories of open generalised polygons, of Hrushovski on approximate equivalence relations and their stabilisers, and of Malliaris and Shelah on new simple theories obtained from hypergraphs.

Groups interpretable in uncountably categorical structures, which are necessarily of finite Morley rank, play an important role in Zilber's structure theory. He obtained some of the foundational results on such groups, and formulated a form of what has become known as the *algebraicity conjecture*, or the *Cherlin–Zilber conjecture*. Poizat's article in this volume discusses its origins in detail, starting with Zilber's

1977 article on categorical groups and rings, and leading to its commonly accepted modern formulation that any simple group of finite Morley rank is an algebraic group over an algebraically closed field. This has been the impetus for a substantial body of work. Along with Poizat's history of the problem, Cherlin's article is devoted to the case of odd type, where much is known, while Borovik's and Deloro's articles concern related results on linearisation of group actions in finite Morley rank.

Later, Boris returned to categoricity in a setting where it is hard to imagine anyone else daring to look for it: the complex exponential field. He developed his *quasiminimality conjecture* that every definable subset is countable or cocountable, and his *pseudoexponentiation conjecture*, which would extend this to a clear account of the coarse geometry of the structure. These, along with the related *conjecture on intersections with tori*, now known in generalised form as the *Zilber–Pink conjecture*, formed the main focus of Boris and his research group in Oxford for many years. The articles of Aslanyan and of Baldwin and Villaveces review various aspects of this programme and its offshoots. Meanwhile, Pila's article proves Zilber–Pink for raising to the power i, Kowalski's examines positive characteristic analogues of the Ax–Schanuel theorem, and Wilkie's article sets out his strategy for an analytic proof of the quasiminimality conjecture.

The above account skips over much of Boris's work. In particular, his thoughts in recent decades have increasingly turned towards applying model-theoretic methods to elucidate some difficulties in mathematical physics. This started with the noncommutative algebras arising in Zariski geometries, and continues through various incarnations of pseudofiniteness and notions of approximation. This work is proving to be a challenge for the model-theoretic community to absorb, perhaps reflected in the fact that, unfortunately, none of the contributions published in this volume ended up being on this subject.

Breaking up these scientific texts, we have included some more historical articles by Hodges, Makowsky, and Yomdin, relating their personal connections to Boris, and their view of the influence of his mathematics. In a similar vein, anecdotes from many people are collected in a piece on Logic Tea in Oxford. Apart from these articles of a biographical or anecdotal character, all articles were refereed to the journal's high standards.

A word on how this volume came to be. When we began thinking how to mark Boris's 75th birthday, the world had closed in response to the Covid-19 pandemic and bore no promise of reopening any time soon. With no clear prospect of an international conference being feasible in 2024, we settled for a "festlose Schrift", a Festschrift without the Fest. As it transpired, we have subsequently been able to arrange a conference, so at the time of writing we are looking forward to presenting the completed volume to Boris in Oxford in September 2024.

We would like to thank all the contributors to this volume for their wonderful articles, and also for being cooperative in making the editorial process a pleasurable task for us. We also thank the journal Model Theory for agreeing to publish this volume, and Martin Hils for guiding the process in his role as chief editor. Finally, we would like to thank Boris himself for his tremendous inspiration and his positive influence on our mathematics, our careers, and on our lives in general.

The Editors

MARTIN BAYS:
bays@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

MISHA GAVRILOVICH:
mishap@sdf.org
University of Haifa, Haifa, Israel

JONATHAN KIRBY:
jonathan.kirby@uea.ac.uk
School of Mathematics, University of East Anglia, Norwich, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.203

Meeting Boris Zilber

Wilfrid Hodges

1. Model theory before Boris

To provide a context, let me say something about the state of model theory when Boris Zilber came into the field in the early 1970s.

An essay of C. C. Chang [1974] entitled "Model theory 1945–1971" should, at least from its title, be ideally suited to telling us what model theory was before Boris Zilber. Alfred Tarski [1954] had proposed the name "model theory" in 1954, on the basis of developments that had started to come together in the decade or so before; Chang's choice of 1945 makes a very reasonable start date. At the other end, Chang mentions some twenty papers dated 1972. This is still too early to include Boris; his earliest publication seems to be in 1974, though he himself cites unpublished papers of his from 1972 and 1973.

Chang divides up model theory into a few dozen "nodes", nearly all of which are either theorems or definitions. He draws a diagram to indicate which nodes were influenced by earlier nodes within model theory. Thus he has four "root" nodes which influence other nodes but are not themselves influenced by other nodes; for example the node "Löwenheim–Skolem–Tarski theorems". A fifth node "Omitting and realizing types (over a set A)" could have been counted as a root, but Chang sees it as influenced by several other nodes in complicated ways. This node is listed as influencing the following node among others:

(1) The notions of stability, rank, degree, finite cover property, etc.; categoricity theorems [Morley 1965; 1967; Baldwin and Lachlan 1971; Shelah 1971].

Chang doesn't reckon that this node (1) influenced any others. Probably if Chang had continued the diagram to include Boris, Boris would have been either in (1) or in a new node "influenced by" (1).

One major influence on Boris from this period is mentioned only indirectly by Chang; this is the collection of questions in circulation about theories categorical in some power. These questions include:

(2) If a countable first-order theory is λ-categorical for some uncountable λ, then must it be λ-categorical for every uncountable λ? (Łoś [1954] stated this as an

MSC2020: primary 03C45; secondary 03-03.
Keywords: history.

upward question plus a downward question. Morley [1965] proved it in both directions.)

(3) Is there a theory in a countable first-order language which is λ-categorical for every uncountable cardinal λ, but not ω-categorical, and is finitely axiomatisable? (Stated in the last section of [Morley 1965], but Morley says it is not his question. Mikhail Peretyat'kin proved an affirmative answer by methods very different from Boris's.)

(4) Is there a theory in a countable first-order language which is λ-categorical for every transfinite cardinal λ and is finitely axiomatisable? (I don't know who first stated this question, but it is a natural counterpart to (3).)

These and other questions of the time shared an important feature. The relevant first-order theories include those of some well-studied classical structures such as algebraically closed fields or vector spaces over finite fields. So it is natural to ask whether we can generalise from the classical structures to a class of structures defined by first-order theories. When Boris came into model theory, Morley had proved an affirmative answer to the question (2) by generalising transcendence rank to Morley rank, and Boris himself would later do something similar to prove a negative answer to (4). But Chang has no node that naturally covers arguments of this kind or the questions that generate them. For example he "exclude[s] from our consideration… model theory applied to algebra, analysis, and set theory" (p. 173). With hindsight I think we have to say that this marked a blind spot in Chang's picture of model theory. But he was not alone in this.

In a recent online interview Boris puts a related point in his own words:

(5) The essence of model theory is an attempt — speaking in more general or philosophical terms — to interpret mathematics as a whole, analysing the language and the logic of it.… You approach every mathematical area or problem, in number theory, in real or complex analysis, even in physics, and ask what is the adequate language and accordingly adequate formalism for this specific area. It might be that a specific problem requires a specific formalism. Then when you identify this formalism, you can approach it as a study of general patterns of formal theories. [Yeh 2018]

We will come back to this below.

Excuse me if I say a word about how I came into model theory. For my university education I signed up for a Hastings Rashdall scholarship at New College Oxford. This scholarship was intended to train future theologians by teaching them Latin, Greek, Greek and Roman history and some modern philosophy. The scholarship converted me to atheism, which didn't fit my proposed career. On the advice of the philosopher Gilbert Ryle (advice which I later learned had been crafted by my philosophy tutor David Wiggins — let me thank them both for this) I applied to work for a doctorate in Literae Humaniores at Oxford, naming logic and with the intention of getting as far into maths as I could manage self-taught. There was some basis for this: for example, Richard Rado (who was a family friend) gave me some offprints on partition calculus. I joined John Crossley's recently formed research group in mathematical logic, and at an early meeting I met Saunders Mac Lane, who encouraged me to learn about vector spaces. The excellent lectures of two fellow students (John Bell and Alan Slomson) ensured I would call myself a model theorist, and I spent the rest of my career racing — with intermittent success — to catch up with the required mathematics.

2. Boris appears

I believe the first time I was aware of Boris was in March 1975. Paul Henrard had organised a week of model theory at Louvain-la-Neuve. Shelah was the main speaker, and he gave several lectures on "The lazy model-theoretician's guide to stability". They were written up by Greg Cherlin, Janos Makowsky, Alex Wilkie and me, and published as [Shelah 1977] in the volume [Henrard 1977]. On page 17, Shelah writes:

> There exist papers of Zilberg [sic] on $\aleph_1$-categoricity of rings, partially overlapping Cherlin and Reineke.... [Added in Proof June 76: Zilberg [sic] also proved independently that $\aleph_1$-categorical division rings are fields.]

This was the first time I became aware of logicians in Russia who were working in categoricity theory. (Oxford in 1976 hosted some Russian logicians at a meeting on word problems in algebra [Adian et al. 1980], and several model theorists were present including Shelah and Yuri Gurevich. Boris was not and I don't recall that he was mentioned.)

Soon after Henrard's conference the British Broadcasting Corporation announced a series of lessons in Russian, and my wife Helen and I signed up to learn. This was partly in the expectation that I would soon meet "Zilberg" or at least some of his papers. (The first Russian sentence that the BBC taught us was *kran ne rabotayet* ("The tap or faucet doesn't work"); several Russian friends assured me that the sentence was absolutely true.) As soon as I could start to read Russian, I looked up the Russian mathematical journals available in the library at Queen Mary University of London, and found several issues of *Matematicheskiye Zametki* with papers by Boris or his colleague Oleg Belegradek. Queen Mary also had *Algebra i Logika* with papers by Palyutin, Erimbetov and others, though these papers were available in translation at other London libraries, Quite soon manuscripts from Boris did arrive in the West. Apparently Boris heard of a grant that would allow him to visit Wrocław in Poland in 1979–80, and he took it up in order to finalise his work on the third question (4) above. His fullest account at that date [Zilber 1980] appeared in the Proceedings of a conference in Karpacz. Soon afterwards a Polish model theorist came through Boulder, where I happened to be visiting. (I am fairly sure this was Leszek Pacholski, who was the editor of the Karpacz Proceedings. My apologies if this is wrong.) He and I made copies of Boris's paper and sent them to a few dozen model theorists. Unfortunately in writing up Boris had missed a required condition on a polynomial, with the result that he had written a paper in which he allowed division by zero. Very soon both Cherlin and some people in Paris pointed out the gap. Alerted to this, Boris went back to what he had learned in Poland, and on that basis he wrote a short note [Zilber 1981] correcting the error. Meanwhile Cherlin and C. Mills had independently seen how to plug the gap by using the classification of finite simple groups, a point noticed later in the 1980s by several group theorists. Boris's corrected proof made no use of this classification.

The next paper of Boris that I saw was an essay that he had archived with the large Russian database VINITI in Kemerovo in 1977. This essay contained preparatory material for the paper [Zilber 1980] mentioned above. Cherlin had the essay and at the Logic Year in Jerusalem in 1981 he gave copies to some Russian-reading participants. I took a copy back to London, and for the next term I met my student Simon Thomas for coffee each Friday morning to dictate a translation of the essay to him. Simon took thorough notes and made an edited version of them. We sent a copy to Boris in Kemerovo. Later we learned that Boris sent it on to Sasha Borovik in Omsk with a note that by reading it Sasha could simultaneously find out what Boris was working on and learn some English. Unfortunately the top sheet naming Boris went missing, and thus it happened that the first published notice in Russian of Boris's work on groups with finite Morley rank was a summary by Sasha attributing the work to Simon Thomas. Another copy of Simon's writeup found its way to Ali Nesin and influenced his work on simple ω-stable groups. Meanwhile Simon himself rapidly completed an elegant doctoral thesis on classification of simple locally finite groups [Thomas 1983]. The thesis was strictly algebra and not model theory — it came to light that Gary Shute at Michigan State had independently reached the same results within algebra. But I think it is fair to cite Simon's thesis as an example of a convergence of interests between model theorists (as Boris) and specialists in algebraic groups (as Sasha).

Another of Boris's early papers that we in England translated into English was his [Zilber 1991], from a Russian original. David Evans was one of the translators, and it's worth a comment that David knew no Russian (at least at that date — he may have learned some since). But he had a good knowledge of the algebraic background that Boris was assuming. This allowed him to reconstruct Boris's argument from my patchy English translation, and in several places to correct the translation to fit the mathematics. This impressed me as an example of mathematicians communicating through the mathematics itself rather than through a Russian or English text.

3. Meeting Boris in person

During the 1980s the iron curtain still divided Europe. East German logicians were increasingly frustrated at not being allowed any contact with their colleagues in the West. Then Ingo Dahn and Helmut Wolter in East Berlin, specialists in the model theory of fields with exponentiation, discovered a way in which East German logicians could host conferences to which non-German logicians were invited. Thus the Easter Conferences on Model Theory came into being; the first was in 1983 and the last in 1991. The German Democratic Republic ("East Germany" for short) ceased to exist on 3 October 1990, and henceforth there was just one Germany. Easter Conferences before 1991 were sometimes held in the conference centres of East German trade unions. But 1991 was different: we used a STASI training centre, where some of the STASI staff had been allowed to stay on as managers of the training centre, provided that they retrained as staff for the new uses of the building. Nothing to do with logic, but it was an extraordinary experience being served in the restaurant by scrupulously polite ex-STASI staff.

From 1983 onwards, the Easter Conferences had started to bring together logicians from both the Eastern and the Western blocs in Europe — excluding only the West Germans. The Russians were slow to join, but in 1986 Sergei Goncharov came from Siberia. In 1987 Boris came, with his colleague Oleg Belegradek. We met in those eerie halls below Friedrichstrasse Station, where travellers passed between East and West Germany under close inspection by the East German guards and their dogs. The conference went well and was the first of two Easter Conferences that Boris attended.

In the mid 1980s the group theorist Otto Kegel proposed to me that we should organise a Durham Symposium in *Model Theory and Groups*. Since Otto was based in Germany, Peter Neumann joined us as a second British organiser. The symposium took place on 18–28 July 1988 with seventy-five participants. Boris was the central speaker and he gave four lectures on "Finite homogeneous geometries 1–4". Another memorable lecture connected with Boris's work was the announcement by Ehud Hrushovski of his counterexamples to Boris's trichotomy conjecture, using a highly ingenious adaptation of Fraïssé's limit construction. From the discussion at the end of Ehud's lecture, and remarks of Boris elsewhere, I came away with the impression that Boris didn't really have a precise trichotomy conjecture. Rather his view was that some form of trichotomy was to be expected as a classical property, and that on general principle it should follow that some natural abstract (in Boris's words, "logically perfect") conditions could be found under which trichotomy was provable. Ehud's result showed that the categoricity conditions that Boris had used so far were not sufficient. A few years later Ehud and Boris published their joint paper [Hrushovski and Zilber 1993] proving the trichotomy conjecture for "Zariski structures", which added an axiomatisation of a Zariski topology as a further condition. Other sufficient conditions have been found. Meanwhile Ehud's new construction has turned out to be extraordinarily versatile for generating interesting structures.

4. Boris in Kemerovo

In the 1990s Boris and his wife Tamara kindly invited me to their apartment in Kemerovo in Siberia. One of the first items to be explained here was Boris's telephone and its role in Russian history. When Boris Yeltsin was planning his coup, it was important that he could rely on the support of various groups, among them the coal miners in the Kuzbas coal fields, which formed the main industry supporting the town of Kemerovo. In order not to leak his plans to his political rivals, Yeltsin had to use contacts via private telephones. Tamara was a journalist with links to the Kuzbas miners, and so it happened that when Yeltsin was ready to move, Boris's telephone carried the message that Yeltsin could rely on the miners of Kuzbas.

One of the few facts about Kemerovo that did reach the British press was that a man in Kemerovo had killed several people and made them into meat pies which he sold at the Kemerovo railway station. Tamara confirmed to me that there was such a man, and told me that she had visited this man in prison in hopes of learning what had driven him to these actions. But she was too discreet to tell me what if anything she had learned from him.

Boris told me of one morning when his five-year-old son opened their front door and found a dead body on the stairs outside. Kemerovo was at times quite a disorderly town. Anti-Jewish attitudes were not uncommon in Kemerovo — or indeed in other places in Russia. Tamara told me of an occasion when she and Boris had been sitting several rows apart in a crowded bus. The woman sitting next to her launched into a fierce attack on Jews, and Tamara could see that the woman was staring at Boris as she spoke.

In the West we were broadly aware of this situation, and we tended to assume that Boris would want to come to the West as soon as he could. But at first it didn't happen. Boris explained that his father knew no Western languages at all; it would be unconscionable to abandon his father in Siberia, and cruel to bring him to a Western country where he couldn't communicate with anybody. But then in 1999 Oxford University invited Boris to apply for the Professorship in Mathematical Logic in succession to Dana Scott and Angus Macintyre. I was told this professorship was first proposed by the Oxford philosophers, who wanted the teaching of logic in Oxford to be under the guidance of a professor who was expert in mathematical logic but also aware of the needs of philosophers in that area. The philosophy faculty was housed in Merton College, and accordingly Merton College had pride of place on the electoral board for the professorship. When Boris was elected, the warden of Merton College was the sinologist Dame Jessica Rawson; she conscientiously took her duties to include helping to bring over and settle Boris's father. No doubt there were other factors, but this was certainly one of them.

5. How mathematicians communicate?

Some events took place that I heard about partly from Oleg Belegradek and partly from Boris himself. The two accounts are compatible but interestingly different. Oleg told me that Russian universities have a set of necessary and sufficient conditions for a lecturer to be raised to the rank of professor. One of these was that the person concerned must have published a book in the relevant discipline. This was the only condition that Boris failed to satisfy. So Oleg said to Boris "Let me write up for publication the notes of the course that you teach" (naming a course), "and we can arrange to get them published".

The next part I heard from Boris sometime later. Boris told me that he had accepted Oleg's offer and Oleg had given him the write-up for him to check. But in Boris's view, Oleg's teaching style was too formalistic and included an unhelpful amount of detail. So Boris went through Oleg's volume and struck out with a red pen maybe a third of the text, adding nothing.

I heard the next step from Oleg. Boris had returned the write-up to Oleg with extensive deletions marked in red pen. After looking over the deletions Oleg had decided to ignore them, and he sent to the printer a copy of the volume as it was before Boris's deletions. All went well and Boris became professor.

There is a point in telling this story. It's agreed that different mathematicians can have widely different strategies for constructing proofs. For example some mathematicians are happiest if they can construct the proof like a logical deduction, adding line by line to what has already been deduced, until the required theorem emerges as the last line. Others prefer to construct the whole proof in a vague or intuitive form and then fill in the details. (If I understood him correctly, Saharon Shelah once told me that in his experience the first sort of mathematician is unlikely to make mistakes, but the first sort is also more likely to meet questions which he or she will never be able to answer.) Obviously there are other dichotomies between different styles of mathematical thinking.

Boris once told me that there was a first time when he knew that another logician understood what he (Boris) was trying to do. This was when Ehud Hrushovski sent him a preprint in which Ehud cited Bézout's theorem in a context not obviously within algebraic geometry. Boris had allowed Bézout's theorem to guide his thinking in a similar context, but he didn't suppose other logicians would understand this and so he hid Bézout's theorem behind an argument with Morley rank.

How does this relate to Boris's remark quoted at (5) above? Both Boris and Ehud were opening up a new area of research which borrowed a picture (though not apparently formal details) from Bézout's theorem. Were they both asking, from their own points of view, what is "the adequate language and accordingly adequate formalism for this specific area"? I leave this as a question, for fear of fabricating what Boris and Ehud were thinking.

But in any case I came to realise that my gappy mathematical education would never equip me to keep up with recent developments in model theory. So as retirement came into view I chose to move across to the history of logic in Arabic in the middle ages, particularly the logic of Ibn Sīnā (Avicenna); this was connected more directly to my undergraduate training. After working for a couple of decades in this historical field, it strikes me as uncanny how many Boris-like features Ibn Sīnā's logical thinking had. One was the constant pressure to identify and codify features of logical thinking that had not previously been noticed. Another was Ibn Sīnā's decision, apparently in his late teens, to abandon large parts of Aristotle's modal logic and replace them by a new logic based on an "adequate language and accordingly adequate formalism for this specific area". Incidentally both Boris and Ibn Sīnā were born in Uzbekistan (Boris in Tashkent, Ibn Sīnā near Bukhara).

Time rolls on and I travel less. I might not see Boris again. But Helen and I were hugely pleased to meet Boris and Tamara again at the conference *Logical Perspectives* 2018 in St Petersburg. Boris is one of those valuable individuals who enrich not only the lives of the people they meet, but also in a more abstract way the world itself and its culture.

Acknowledgments

It should be obvious that I am in debt to dozens of model theorists for the contents of this paper. I particularly thank the editors of this volume, the referees of the paper and Boris himself for their helpful comments and information.

References

[Adian et al. 1980] S. I. Adian, W. W. Boone, and G. Higman (editors), *Word problems*, *II*: *The Oxford book*, Studies in Logic and the Foundations of Mathematics **95**, North-Holland Publishing Co., Amsterdam, 1980. MR Zbl

[Baldwin and Lachlan 1971] J. T. Baldwin and A. H. Lachlan, "On strongly minimal sets", *J. Symbolic Logic* **36** (1971), 79–96. MR Zbl

[Chang 1974] C. C. Chang, "Model theory 1945–1971", pp. 173–186 in *Proceedings of the Tarski Symposium* (Berkeley, CA, 1971), edited by L. Henkin et al., Proc. Sympos. Pure Math. **XXV**, Amer. Math. Soc., Providence, RI, 1974. MR Zbl

[Henrard 1977] P. Henrard (editor), *Six days of model theory* (Louvain-la-Neuve, Belgium, 1975), Éditions Paul Castella, Albeuve, 1977. MR Zbl

[Hrushovski and Zilber 1993] E. Hrushovski and B. Zilber, "Zariski geometries", *Bull. Amer. Math. Soc.* (*N.S.*) **28**:2 (1993), 315–323. MR Zbl

[Łoś 1954] J. Łoś, "On the categoricity in power of elementary deductive systems and some related problems", *Colloq. Math.* **3** (1954), 58–62. MR Zbl

[Morley 1965] M. Morley, "Categoricity in power", *Trans. Amer. Math. Soc.* **114** (1965), 514–538. MR Zbl

[Morley 1967] M. Morley, "Countable models of $\aleph_1$-categorical theories", *Israel J. Math.* **5** (1967), 65–72. MR Zbl

[Shelah 1971] S. Shelah, "Stability, the f.c.p., and superstability; model theoretic properties of formulas in first order theory", *Ann. Math. Logic* **3**:3 (1971), 271–362. MR Zbl

[Shelah 1977] S. Shelah, "The lazy model-theoretician's guide to stability", pp. 9–76 in *Six days of model theory* (Louvain-la-Neuve, Belgium, 1975), edited by P. Henrard, Éditions Paul Castella, Albeuve, 1977. MR Zbl

[Tarski 1954] A. Tarski, "Contributions to the theory of models, I", *Nederl. Akad. Wetensch. Proc. Ser. A* **57** (1954), 572–581. MR Zbl

[Thomas 1983] S. R. Thomas, *Classification theory of simple locally finite groups*, Ph.D. thesis, Bedford College, University of London, 1983, available at https://www.proquest.com/docview/1791501281. MR Zbl

[Yeh 2018] T. Yeh, "Boris Zilber", online interview, 2018, available at https://www.merton.ox.ac.uk/sites/default/files/inline-files/Boris-Zilber.pdf. Retrieved 1 Jan 2023.

[Zilber 1980] B. I. Zilber, "Totally categorical theories: structural properties and the nonfinite axiomatizability", pp. 381–410 in *Model theory of algebra and arithmetic* (Karpacz, Poland, 1979), edited by L. Pacholski et al., Lecture Notes in Math. **834**, Springer, 1980. MR Zbl

[Zilber 1981] B. I. Zilber, "Solution of the problem of finite axiomatizability for theories that are categorical in all infinite powers", pp. 69–74 in *Investigations in theoretical programming*, edited by M. A. Taĭtslin, Kazakh. Gos. Univ., Alma Ata, 1981. MR Zbl

[Zilber 1991] B. I. Zilber, "Groups and rings whose theory is categorical", pp. 1–16 in *Thirteen papers in algebra*, *functional analysis*, *topology*, *and probability*, *translated from the Russian*, edited by S. Ivanov, Amer. Math. Soc. Transl. Ser. 2 **149**, 1991. Translated by D. Evans and W. Hodges from *Fund. Math.* **95**:2 (1977), 173–188. MR Zbl

Received 24 Jan 2023. Revised 22 Mar 2023.

WILFRID HODGES:

wilfrid.hodges@btinternet.com
Herons Brook, Sticklepath, Okehampton, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.213

Very ampleness in strongly minimal sets

Benjamin Castle and Assaf Hasson

Dedicated to Boris Zilber, whose far-reaching vision is a never ending source of inspiration.

Inspired by very ampleness of Zariski geometries, we introduce and study the notion of a very ample family of plane curves in any strongly minimal set and the corresponding notion of a very ample strongly minimal set (characterized by the definability of such a family). We show various basic properties; for example, any strongly minimal set internal to an expansion of an algebraically closed field is very ample, and any very ample strongly minimal set nonorthogonal to a strongly minimal set Y is internal to Y. We then use very ampleness to characterize the *full relics* of an algebraically closed field K — those structures $\mathcal{M} = (M, \dots)$ interpreted in K which recover all constructible subsets of powers of M. Next we show that very ample strongly minimal sets admit very ample families of plane curves of all dimensions, and we use this to characterize very ampleness in terms of definable pseudoplanes. Finally, we show that nonlocally modular expansions of divisible strongly minimal groups are very ample, and we deduce — answering an old question of Martin (1988) — that in a pure algebraically closed field K there are no reducts between $(K, +, \cdot)$ and $(K, \cdot)$.

1.	Introduction	214
2.	Notation and preliminaries	217
3.	Very ampleness: first properties	218
4.	Applications of very ampleness	230
5.	Technical results on strongly minimal structures	239
6.	Very ample families of a prescribed dimension	243
7.	Very ampleness in strongly minimal groups	251
	Acknowledgement	257
	References	257

Both authors were partially supported by ISF grant No. 555/21. Castle was supported by a Ben Gurion Kreitman fellowship.
MSC2020: 03C45.
Keywords: strongly minimal sets, ACF relics, ACF reducts, nonlocal modularity, Zilber trichotomy.

1. Introduction

Many ideas in stability theory (such as Zilber's classification of totally categorical structures, see, e.g., [Zilber 1993]) build heavily on or are inspired by Zilber's weak trichotomy theorem, stating that any strongly minimal set is either locally modular or defines a rank 2 pseudoplane (recall that a *pseudoplane* is an abstract incidence relation $I \subset P \times L$ between "points" and "lines" such that each point lies on infinitely many lines, each line contains infinitely many points, any two lines have finite intersection, and any two points lie on only finitely many common lines). Indeed, Zilber's (strong) trichotomy conjecture, in one of its many variants, asserted that every uncountably categorical pseudoplane should be mutually interpretable with an algebraically closed field.

While Zilber's conjecture is false [Hrushovski 1993], it has been shown to hold in many instances. For example, Castle [2023] has recently shown that the conjecture holds assuming the pseudoplane is itself interpreted in an algebraically closed field of characteristic zero, and Hasson and Sustretov [2017] have shown the same for rank preserving interpretations in all characteristics.

Contexts where Zilber's trichotomy is known to hold are of a geometric nature, and the geometry plays a major part in many proofs of restricted versions of the trichotomy. Most such proofs, given a nonlocally modular strongly minimal set S, make use of almost faithful families of plane curves (families of one-dimensional subsets of S^2 with few infinite intersections), rather than pseudoplanes. The reason is that, although such objects are combinatorially harder to work with than definable pseudoplanes, the good geometric properties of S can be exploited to study and control them. And while S does always interpret a pseudoplane, the set of "points" of the pseudoplane might be imaginary, making it much harder (if at all possible) to use the geometric properties of S to construct a field.

However — in the above notation — the existence of a definable pseudoplane in S^2 could simplify the construction of the field in many instances. For example, it seems that Rabinovich's work [1993] would be much simpler if one assumes the set of points of the given pseudoplane to be the affine plane[1]. The lack of a pseudoplane on S^2 is also at the technical heart of [Castle 2023], and moreover explains a key reason that strongly minimal expansions of groups are historically easier to handle — since in the case of groups, one can always find such a pseudoplane (potentially after quotienting by a finite subgroup).

A crucial difference between a definable pseudoplane in S^2 and a 2-dimensional family of plane curves is that the latter allows the existence of *semi-indistinguishable points*. These are certain exceptional pairs of points incident to infinitely many

[1]See the discussion on p. 2–3 of [Rabinovich 1993]. Unfortunately, the manuscript [9] that is referred to in that text does not seem to be available.

common curves, which underlay many combinatorial intricacies of trichotomy proofs. Semi-indistinguishable points tend to prohibit the "full" recovery of the underlying geometry. Indeed, already in [Martin 1988], it was pointed out (in an example attributed to Hrushovski, see Example 3.6) that a reduct of an algebraically closed field need not be constructible (i.e., quantifier free definable in the language of rings) in any copy of the field it interprets (or even $\mathcal{M}$-definably isomorphic to such a constructible set).

In their seminal work, Hrushovski and Zilber [1996] identified a condition they call *very ampleness*, assuring that a Zariski Geometry, $\mathcal{Z}$, is algebraic in the above sense, i.e., that not only does it interpret a field, K, but it is definably isomorphic (as a Zariski geometry) to a smooth K-algebraic curve. Very ampleness (as defined by Hrushovksi and Zilber) does not go so far as to require a pseudoplane on the set Z^2, but it is a related notion — namely, it asserts the existence of a family of plane curves that separates points in Z^2 in an obvious sense.

In the present paper, we introduce a similar notion of very ampleness[2] for arbitrary strongly minimal sets, characterized by the presence of families of plane curves with no semi-indistinguishable pairs (Corollary 3.21). In fact, we show that the presence of one such family implies the existence of such families of arbitrary dimension (Corollary 6.5); using this, we conclude that a strongly minimal set D is very ample if and only if there is a definable pseudoplane whose points are (a generic subset of) D^2 (Proposition 6.7). Moreover, we show that in the context of Zariski geometries, our version of very ampleness is a necessary and sufficient condition for a Zariski geometry to be algebraic, and is thus equivalent to the original notion (Proposition 4.8).

We apply very ampleness to the study of structures interpretable in algebraically closed fields. We show that every algebraic curve over such a field is very ample (Corollary 3.29). We then show that a very ample strongly minimal set is internal to every strongly minimal set it interprets, a phenomenon allowing for the identification of very ample ACF-interpreted structures with algebraic curves (Proposition 4.1). In particular (noting the known instances of the Zilber trichotomy in the relevant structures, see [Castle 2023; Hasson and Sustretov 2017]), we conclude with an analogue of the main theorem on Zariski geometries (see Theorem 4.14):

Theorem 1. *Let K be an algebraically closed field, M be constructible over K, and $\mathcal{M} = (M, \dots)$ be a strongly minimal reduct of the full K-induced structure on M. Assume $\mathcal{M}$ satisfies the Zilber trichotomy*[3]*. Then the following are equivalent*:

(1) *$\mathcal{M}$ is very ample.*

[2] In [Hrushovski and Zilber 1996] the weaker notion of *ampleness*, equivalent to nonlocal modularity, is used. We do not use this term here.

[3] By "[A structure] $\mathcal{M}$ satisfies Zilber's trichotomy" we mean that every nonlocally modular strongly minimal set definable in $\mathcal{M}$ interprets an algebraically closed field. For ACF-relics, this assumption is redundant, since Zilber's trichotomy is true for those. But as a complete proof has not,

(2) *$\mathcal{M}$ is isomorphic, outside of a finite set, to the full K-induced structure on some irreducible algebraic curve over K.*

(3) *Every constructible subset of every power of M is definable in $\mathcal{M}$.*

For arbitrary ACF-interpreted structures, we get the next result (see Theorem 4.16).

Theorem 2. *Let K be an algebraically closed field, and assume the Zilber trichotomy holds for strongly minimal structures interpreted in K. Let M be constructible over K, and let $\mathcal{M} = (M, \dots)$ be an arbitrary reduct of the K-induced structure on M. Then the following are equivalent:*

(1) *$\mathcal{M}$ is almost strongly minimal, and every strongly minimal set in $\mathcal{M}$ is very ample.*

(2) *Every constructible subset of every power of M is definable in $\mathcal{M}$.*

We then apply Theorem 1 in the case of groups, recovering, in particular, the following (see Theorem 7.8):

Theorem 3. *Let K be an algebraically closed field, and let $(G, \cdot)$ be a one-dimensional divisible algebraic group over K. Let $\mathcal{G}^{\mathrm{Zar}}$ be the full K-induced structure on G, and let $\mathcal{G}^{\mathrm{lin}}$ be the structure endowing G with the group operation and all K-definable endomorphisms of G. Then there are no intermediate structures between $\mathcal{G}^{\mathrm{lin}}$ and $\mathcal{G}^{\mathrm{Zar}}$.*

This expands the main result in [Marker and Pillay 1990] and can be seen as an algebraically closed field analogue of a recent similar result of Abu Saleh and Peterzil [2023] for real closed fields.

For strongly minimal groups, divisibility is equivalent to unbounded exponent. In the case of finite exponent, the statement of Theorem 3 fails: in positive characteristic, Marker and Pillay [1990] give an expansion of $(K, +)^{\mathrm{lin}}$ that interprets a field but does not define multiplication (see the discussion after Corollary 1.8 *loc. cit.*).

Along the way, we prove several technical results that may be of interest on their own right. We single out the following (Theorem 7.3):

Theorem 4. *If $\mathcal{G}$ is a strongly minimal expansion of a group, then $\mathcal{G}$ is not locally modular if and only if there exists a definable $X \subseteq G^2$ that is not a finite boolean combination of cosets of definable subgroups of G.*

This result, while well known among experts, does not seem to exist in writing in full generality, and since we needed it in the present work we took the opportunity to give the details.

The paper is written with an eye toward the proofs of Theorems 1–4 above, but also as a possible reference for future work around Zilber's trichotomy and other

at the time of the writing of this paper, appeared in print, so we chose a formulation of the results whose correctness is independent of Zilber's restricted Trichotomy questions related to the fine structure of strongly minimal sets and the structures they interpret. For that reason, some of the proofs are stated and proved in somewhat greater generality than is actually needed. Thus, for example, the existence of arbitrarily large very ample families of plane curves (Proposition 6.3) is not explicitly used in the text, but significant parts of its proof (e.g., very ampleness of algebraic curves, see Corollary 3.29) are essential for our arguments.

2. Notation and preliminaries

Throughout, we work in a saturated enough model of a stable theory. Except for the general definition of very ample types (invoked only in Corollary 6.12) the work could be carried out in any structure where all definable strongly minimal sets are stably embedded. Since stable embeddedness is equivalent to uniform stable embeddedness, given a strongly minimal set X there is no harm in assuming that all definable families of subsets of X^n we are considering are parameterized by X-definable sets.

We are using standard model theoretic terminology. For basic concepts such as Morley rank, strong minimality, local modularity, one-based theories, etc., we refer the reader to any textbook covering the first chapters of geometric stability theory such as [Marker 2002], [Pillay 1996], or [Tent and Ziegler 2012].

After fixing a structure $\mathcal{M}$, unless otherwise stated, the word *definable* refers to definability in $\mathcal{M}^{\mathrm{eq}}$ with parameters. All parameter sets are smaller than the level of saturation of $\mathcal{M}$, and are usually denoted, A, B. We use the standard model theoretic abuse of notation and write $a \in M$ instead of $a \in M^{|a|}$, allowing a to also be an element of an imaginary sort.

If X is a strongly minimal definable set, the notation X^{eq} refers to the union of all sorts formed by quotienting definable subsets of powers of X by definable equivalence relations. A definable set in X^{eq} is *stationary* if it has Morley degree one — equivalently, if it has a unique generic type over any set of parameters defining it. It follows by uniform stable embeddedness that every stationary set S in X^{eq} has a canonical base in X^{eq} (i.e., the canonical base of its generic type), which we denote $\mathrm{Cb}(S)$. We use dim and *dimension* to denote Morley rank of definable sets in X^{eq}. If $Y \subseteq Z$ are definable sets in X^{eq}, we say that S is large in Z if $\dim(Z \setminus Y) < \dim(Z)$ and Y is small in Z if $\dim(Y) < \dim(Z)$. A generic subset of Z is a definable subset of Z that is not small. Thus, if $Y \subset Z$ and Z is stationary, Y is generic in Z if and only if it is large in Z.

We will use implicitly the definability of Morley rank in X^{eq} for strongly minimal X; see [Baldwin 1973]. We will also use the resulting fact that, given a small parameter set A, a definable set C in X^{eq} is always a generic member of an A-definable family of $\mathcal{M}$-definable subsets of the same Morley rank.

By a *curve* (in X^{eq}) we mean a one-dimensional definable set, and by a *plane curve* in X, we mean a curve in X^2. Note that we do not require curves to be strongly minimal. A plane curve C is *trivial* if one of its projections has an infinite fiber. If C and D are nontrivial plane curves, then their composition $D \circ C$ is — in analogy with the composition of functions — the curve $\{(x, z) : (\exists y)((x, y) \in C \wedge (y, z)) \in D\}$.

A definable family of plane curves $\mathcal{C} := \{C_t : t \in T\}$ is *faithful* if $t \neq t'$ implies $\dim(C_t \bigtriangleup C'_t) = 0$. In greater generality, a definable family of definable sets in X^{eq} is faithful if the symmetric difference of any two members of the family is small in both. The total space (or the graph) of the family $\mathcal{C}$ is denoted $C := \{(x, t) \in X^2 \times T : x \in C_t\}$.

A definable set C is *almost contained* in a definable set C' if $\dim(C \setminus C) < \dim(C)$. The set C is *almost equal* to C' if C is almost contained in C' and C' is almost contained in C (equivalently, if $\dim(C \bigtriangleup C') < \dim(C) = \dim(C')$). This is a definable equivalence relation (on families of definable sets). It is now a standard and easy exercise to show that if C is any stationary set in X^{eq} with $c := \mathrm{Cb}(X)$, then, up to a set of dimension smaller than $\dim(C)$, the set C is a member of the $\varnothing$-definable family $\varphi(x, y) \wedge \theta(y)$ (parameterized by $\theta(y)$), where $\theta(y) \in \mathrm{tp}(c/\varnothing)$ isolates the type (in its rank and degree) and $\varphi(x, c)$ isolates the generic type of C over c in its rank and degree.

A strongly minimal set X is *locally modular*, if any definable faithful family of plane curves in X^2 is (at most) one-dimensional. By this we mean that if such a family is parameterized by a set T then $\dim(T)$, the dimension of the family, is at most one. We will systematically use (without further reference) the equivalence of local modularity and one-basedness, which — in turn — is equivalent to the fact that $\dim(\mathrm{Cb}(S)) \leq 1$ for any strongly minimal $S \subseteq X^2$ [Pillay 1996, §II Proposition 2.6].

3. Very ampleness: first properties

3A. *Definition and first examples.* We start with a general definition, that we later specialize almost exclusively to the strongly minimal setting:

Definition 3.1. Let p be a stationary type, with canonical base c, and let A be a set of parameters. We say that p is *very ample over A* if for any two distinct realizations $x, y \models p$, x and c fork over Ay.

Definition 3.2. Let X be a strongly minimal set, definable over a set A. A strongly minimal $C \subset X^2$ is *very ample in X over A* if the generic type of C is very ample over A.

Definition 3.3. Let X be a strongly minimal set. We say that X is *very ample* if, for some set A such that X is A-definable, there is a very ample strongly minimal plane curve in X over A.

Definition 3.4. Let $\mathcal{M}$ be a strongly minimal structure. We say that $\mathcal{M}$ is very ample if its universe is very ample as a strongly minimal set.

Example 3.5. Let $(K, +, \cdot)$ be an algebraically closed field, and let $(a, b) \in K^2$ be generic. We claim that the line $L_{(a,b)}$ defined by $y = ax + b$ is very ample in K over $\varnothing$, and thus K is very ample. To see this, first note that the canonical base of the generic type of $L_{(a,b)}$ is just (a, b). Now let $z_1 = (x_1, y_2)$ and $z_2 = (x_2, y_2) \in L_{(a,b)}$ be distinct generics over (a, b). One then easily computes that $\dim(ab/z_2) = 1$, essentially because the family of lines through a point in the plane is one-dimensional. On the other hand, because any two distinct points determine exactly one line, we have $\dim(ab/z_1z_2) = 0$, which gives the desired dependence.

Example 3.6. Suppose $(K, +, \cdot)$ is an algebraically closed field, and consider the reduct $\mathcal{M}$ of K obtained by endowing K with the relations $x^2 + y^2 = z^2$ and $x^2y^2 = z^2$. Then $\mathcal{M}$ is nonlocally modular, but is not very ample. Indeed, suppose $C \subset K^2$ is a strongly minimal plane curve which is very ample over A, and let $c = \mathrm{Cb}(C)$. Let $(x, y) \in C$ be generic over Ac. Without loss of generality, assume x is generic in K over Ac (otherwise y is). Note that any function $\sigma : K \to K$ with the property that $\sigma(x) \in \{x, -x\}$ for all x is an automorphism of $\mathcal{M}$. In particular, there is an automorphism fixing Ac pointwise but sending x to $-x$, and y to some $z = \pm y$. It follows that $\mathrm{tp}(-x, z/Ac) = \mathrm{tp}(x, y/Ac)$, so $(-x, z)$ is generic in C over Ac. Then, by very ampleness, $(-x, z)$ should fork with c over Axy; but this is impossible because $(-x, z)$ is algebraic over (x, y).

Remark 3.7. We note that there is a significant difference between a strongly minimal set, X, being very ample (meaning that it admits a very ample plane curve) and the generic type of X being very ample. As we will see in the sequel, the existence of a (strongly minimal) very ample complete type is equivalent to the structure being non one-based. The former condition is stronger and is equivalent to the existence of a definable pseudoplane whose set of points is generic in X^2. Except for Corollary 6.12 the latter property is reserved for plane curves in X (and we will say that C is very ample in X), so hopefully no confusion will arise.

3B. ***Very ampleness and nonlocal modularity.*** It follows directly from the definition that a very ample strongly minimal set is not locally modular.

Lemma 3.8. *Let X be strongly minimal and definable over A, let C be a strongly minimal plane curve in X, and let $c = \mathrm{Cb}(C)$. If C is very ample in X over A, then $\dim(c/A) \geq 2$. In particular, if X is very ample, then X is not locally modular.*

Proof. Let x and y be independent generics in C over Ac. By very ampleness, c forks with x over Ay, which implies in particular that $\dim(c/Ay) \geq 1$.

Claim 3.8.1. $$\dim(y/A) = 2.$$

Proof. Otherwise, y would belong to a one-dimensional set over A, which must have C as a strongly minimal component. But this would force $c \in \mathrm{acl}(A)$, contradicting $\dim(c/Ay) \geq 1$. □

Now by the claim, and since $\dim(y/Ac) = 1$ by definition, it follows that y forks with c over A. By symmetry this implies

$$\dim(c/A) > \dim(c/Ay) \geq 1,$$

i.e., $\dim(c/A) \geq 2$. $\square$

We will see later on, Corollary 6.12, that the weaker condition of X^{eq} containing a very ample type is equivalent to nonlocal modularity.

Remark 3.9. It also follows from this lemma that if a strongly minimal plane curve C is very ample in X over some set, then C is *nontrivial*, i.e., both projections $C \to X$ are finite-to-one. Indeed, otherwise C would agree up to finitely many points with $\{c\} \times X$ or $X \times \{c\}$ for some $c \in X$, implying that $c = \mathrm{Cb}(X)$ has dimension at most 1 over A.

Note that if X, Y are strongly minimal sets such that $Y \triangle X$ is finite, then for any nontrivial strongly minimal plane curve $C \subseteq X^2$ the set $C \setminus Y^2$ is finite, so $C \cap Y^2$ and C have the same generic type. In particular, if C is very ample in X then $C \cap Y^2$ is very ample in Y. Thus X is very ample if and only if Y is.

Before moving on we mention the following useful and related observation, which may help to further clarify the definition. Roughly, it shows that the non-very ampleness of a nonlocally modular strongly minimal set always arises from an interalgebraicity relation on the plane.

Lemma 3.10. *Let X be strongly minimal and A-definable. Let C be a plane curve in X, definable over $B \supset A$. Assume that for every strongly minimal component S of C we have* $\dim(\mathrm{Cb}(S)/A) \geq 2$. *Let $x, y \in C$ be any two distinct generics over a parameter set $B \supseteq A$. Then either x forks with B over Ay, or x and y are interalgebraic over A.*

Proof. Throughout we assume $A = \varnothing$. Let $S \subset C$ be a strongly minimal component containing x and let $s = \mathrm{Cb}(S)$. Note that $s \in \mathrm{acl}(B)$ and $\dim(x/s) = 1$ (since x is generic in C). Now, by assumption, $\dim(s) \geq 2$; in particular, $s \notin \mathrm{acl}(\varnothing)$. So $\mathrm{tp}(x/s)$ must fork over $\varnothing$, and thus $\dim(x) = 2$. We now argue in two cases:

Case 1: Suppose $x \in \mathrm{acl}(y)$. Since $\dim(x) = 2$, this implies $\dim(y) = 2$. Thus, x and y are interalgebraic, proving the lemma in this case.

Case 2: Suppose $x \notin \mathrm{acl}(y)$; we show that x forks with B over y. By assumption, we have $\dim(x/y) \geq 1$. Note also that $\dim(x/By) \leq 1$ (as $x \in C_t$). Now suppose toward a contradiction that $x \mathop{\smile\!\!\!\!\!\!\;|}_y B$. Then clearly we have $\dim(x/y) = \dim(x/By) = 1$. Since $s \in \mathrm{acl}(B)$, also $\dim(x/Bsy) = 1$. In particular, $\mathrm{tp}(x/Bsy)$ (i.e., the generic type of S over Bsy) does not fork over y; so since $s = \mathrm{Cb}(S)$, we get that $s \in \mathrm{acl}(y)$. Since $y \in X^2$, this gives $\dim(s) \leq 2$. Combined with the assumption that $\dim(s) \leq 2$, we therefore have $\dim(s) = 2$, which implies that s and y are interalgebraic. But then $\dim(y/s) = \dim(y/Bs) = 0$, and since $s \in \mathrm{acl}(B)$ this implies $\dim(y/B) = 0$, contradicting that y is generic in C over B. $\square$

3C. *Extending the definition.* In many applications, it is more convenient to work with families of plane curves that are not necessarily generically strongly minimal. We extend our definition of very ampleness to such (families of) curves. To do this, we observe that the choice of defining parameter for a plane curve, C, does not affect Definition 3.2 in any significant way.

Lemma 3.11. *Let X be a strongly minimal and A-definable set, and let C be an At-definable plane curve in X for some tuple t. The following are equivalent:*

(1) *For any two distinct $x, y \in C$ each generic over At, x forks with t over Ay.*

(2) *For any two distinct $x, y \in C$ each generic over At, $\dim(t/Axy) \leq \dim(t/A) - 2$.*

(3) *For every strongly minimal set $S \subset C$, $\dim(\mathrm{Cb}(S)/A) \geq 2$, and for any two distinct $x, y \in C$ each generic over At, x and y are not interalgebraic over A.*

Proof. Throughout, we assume $A = \varnothing$. By our standing assumption that X is stably embedded, we may assume that t is a tuple in X.

(1) $\Longrightarrow$ (2): Assume (1), and let x and y be distinct generics of C_t over t.

Claim 3.11.1. $$\dim(y) = 2.$$

Proof. Let y' be an independent realization of $\mathrm{tp}(y/t)$ over ty. Then by (1), y forks with t over y'. Since $\dim(y/t) = 1$ by assumption, we obtain $\dim(y/y') = 2$, which implies the claim. □

Now since $\dim(y) = 2$ and $\dim(y/t) = 1$, it follows that $\dim(t/y) = \dim(t) - 1$. By (1) again, t forks with x over y, which gives $\dim(t/xy) \leq \dim(t) - 2$, as desired.

(2) $\Longrightarrow$ (3): Assume (2). First let S be a strongly minimal component of C, with canonical base s. Let x and y be independent generics in S over st. Then $\dim(txy) = \dim(t) + 2$, and by (2) $\dim(t/xy) \leq \dim(t) - 2$, which combined gives $\dim(xy) = 4$. On the other hand $\dim(sxy) = \dim(s) + 2$, so since $\dim(xy) = 4$ we get $\dim(s/xy) = \dim(s) - 2$. In particular, $\dim(s) - 2 \geq 0$, so $\dim(s) \geq 2$.

Now to complete the proof of (3), suppose x and y are any two distinct generics of C over t. Then $\dim(txy) \geq \dim(tx) = \dim(t) + 1$, while by (2) $\dim(t/xy) \leq \dim(t) - 2$. It follows that $\dim(xy) \geq 3$, while if x and y were interalgebraic we would have $\dim(xy) \leq 2$. Thus, we have shown (3).

(3) $\Longrightarrow$ (1): Assume (3). Let x and y be a counterexample to (1). By (3), the hypotheses of Lemma 3.10 are satisfied, and applying the lemma gives that x and y are interalgebraic over A, contradicting (3). □

Corollary 3.12. *Let X be strongly minimal and A-definable. For $i = 1, 2$ let C_i be an At_i-definable plane curve in X for some tuple t_i. Assume that C_1 and C_2 have finite symmetric difference. Then the equivalent conditions of Lemma 3.11 hold for C_1 over At_1 if and only if they hold for C_2 over At_2.*

Proof. We use (3) of the lemma. Since the C_i have finite symmetric difference, the canonical bases of their strongly minimal components coincide. So, by symmetry, it suffices to show that the failure of the second clause of (3) for C_1 and t_1 implies the failure of the same clause for C_2 and t_2. To that end, let x and y be distinct generics in C_1 over At_1, and assume they are interalgebraic over A. Let (x', y') be an independent realization of $\operatorname{tp}(xy/At_1)$ over At_1t_2. Then x' and y' are independent generics over C_2 over At_2, and are also interalgebraic over A, as desired. □

The following is now well defined:

Definition 3.13. Let X be strongly minimal and A-definable. Let C be a plane curve in X. Then C is *very ample in X over A* if any of the equivalent conditions of Lemma 3.11 hold for C over At, for some (equivalently, any) tuple t such that C is At-definable.

It is immediate from the last corollary that Definitions 3.2 and 3.13 coincide for strongly minimal plane curves, since up to a finite set any such curve is definable over its canonical base. Let us also point out the following:

Corollary 3.14. *Let X be strongly minimal and definable over A. Let C be a plane curve in X. If C is very ample in X over A, then so is every strongly minimal component of C.*

Proof. Assume C, and all of its strongly minimal components, are definable over At for some tuple t, and let $S \subset C$ be any such component. Let $x \neq y$ be generics in S over At. Then x and y are also generics of C over At; so by the very ampleness of C over A, x forks with t over Ay, which shows that S is also very ample over A. □

The converse of the last corollary is not true: there can be non-very ample plane curves every component of which is very ample, as in the following example:

Example 3.15. Let $\mathcal{G} = (G, +, \dots)$ be a strongly minimal expansion of a group, and let $S \subset G^2$ be a $\varnothing$-definable strongly minimal plane curve whose generic type has trivial stabilizer (for example $\mathcal{G}$ could be the additive group of the complex field, and S could be the graph of $y = x^2$). Now let $a \in G^2$ be generic, let $S_a := S + a$, and let $C = S_a \cup -S_a$. It is easy to see that for $s \in S_a$ generic over a, the pair $s, -s \in C$ contradicts the very ampleness of C (by (3) of Lemma 3.11). However, each of S_a and $-S_a$ is very ample; see Example 3.18 below.

3D. ***Very ampleness in families.*** The reader may find our definition of very ampleness foreign in light of the analogous notion in [Hrushovski and Zilber 1996] (a precise statement on the relation of our notion of very ampleness and the original term in the context of Zariski geometries requires additional preparation, and can be found in Section 4B). We hope that the present subsection will help to explain for now why our definition captures the original idea of the term. To that end, we now define a notion of very ampleness for families of plane curves.

Notation. If S is definable and $X := \{X_t : t \in T\}$ is a definable family of subsets of S, then for $s \in S$ we denote the set $\{t \in T : s \in X_t\}$ by X^s.

Definition 3.16. Let X be strongly minimal and A-definable, and let $\mathcal{C} = \{C_t : t \in T\}$ be an A-definable family of plane curves in X.

(1) $\mathcal{C}$ is *very ample* if for any $x \neq y \in M^2$, we have $\dim(C^x \cap C^y) \leq \dim T - 2$.

(2) $\mathcal{C}$ is *generically very ample* if for any generic $t \in T$ over A and any distinct $x, y \in C_t$ generic over At, we have $\dim(t/Axy) \leq \dim T - 2$.

Observe that, in particular, a generically very ample family of plane curves is at least two-dimensional.

Example 3.17. Let $(K, +, \cdot)$ be an algebraically closed field. Then, similarly to Example 3.5, one shows easily that the family of lines $y = ax + b$ is very ample, because any two distinct points determine exactly one line.

The following example is well known and easy:

Example 3.18. Let $(G, +)$ be a strongly minimal expansion of a group, and let $C \subset G^2$ be a strongly minimal plane curve. Then the family of translates of C, $\{C + t : t \in G^2\}$, is very ample if and only if the generic type of C has trivial stabilizer. Indeed, $t_1 + C$ is almost equal to $t_2 + C$ if and only if $t_1 - t_2 \in \mathrm{Stab}(p)$ for p the generic type of C.

Proposition 3.19 below, in addition to the ensuing two useful corollaries, shows the relationship between the notions of very ampleness we have defined so far (of a plane curve, of a strongly minimal set, and of a family of plane curves).

Proposition 3.19. *Let X be strongly minimal and A-definable, let $\mathcal{C} = \{C_t : t \in T\}$ be an A-definable family of plane curves in X, and let $C \subset X^2 \times T$ be the graph of $\mathcal{C}$. Then the following are equivalent*:

(1) *$\mathcal{C}$ is generically very ample.*

(2) *There is a nongeneric A-definable set $Z \subset C$ such that $C - Z$ is the graph of a very ample family of plane curves in X.*

(3) *For any generic $t \in T$ over A, C_t is very ample in X over A.*

Proof. The equivalence of (1) and (3) is Lemma 3.11.

(1) $\Longrightarrow$ (2): Let $p(x, y, z)$ be a partial type over A asserting that $x \neq y \in X^2$ and that (x, z) and (y, z) are both generics in C over A. It is a restatement of (1) that for $(x, y, z) \models p$, we have $\dim(z/Axy) \leq \dim(T) - 2$. By compactness, this is witnessed by a finite part of p. We thus obtain a nongeneric $Z \subset C$ such that for $x \neq y$, we have $\dim((C - Z)^x \cap (C - Z)^y) \leq \dim T - 2$. This almost implies (2). Technically, though, one should add to Z all C_t having only finitely many points surviving in $C - Z$ (so that $C - Z$ is a family of plane curves); one should then

verify that Z is still nongeneric in C, and that set of "surviving" $t \in T$ in $C - Z$ is generic in T. Both of these are easy to do.

(2) $\implies$ (1): Let Z be as in (2), and let $p(x, y, z)$ be the same type considered in the case above. We show that for $(x, y, z) \models p$ we have $\dim(z/Axy) \leq \dim T - 2$, which as above is equivalent to (1). So take such (x, y, z). We may assume that $(x, y, z) \mathop{\smile\!\!\!\!\!\!\mid}_A B$ where Z is B-definable; thus (x, y, z) is generic in C over AB, so $(x, y, z) \in C - Z$. Then by very ampleness, $\dim(z/ABxy) \leq \dim(T) - 2$, and again since $(x, y, z) \mathop{\smile\!\!\!\!\!\!\mid}_A B$ this implies $\dim(z/Axy) \leq 2$, as desired. □

To sum up, we show that a plane curve is very ample precisely when it coincides, up to a finite set, with a generic member of a very ample family.

Corollary 3.20. *Let X be strongly minimal and A-definable. Let C be a plane curve in X. Then the following are equivalent*:

(1) *C is very ample in X over A.*
(2) *C has finite symmetric difference with an A-generic member of an A-definable very ample family of plane curves in X.*
(3) *Whenever T is a stationary definable set, $\mathcal{C} = \{C_t : t \in T\}$ is an* $\operatorname{acl}(A)$*-definable family of plane curves in X, and $t \in T$ is a generic element over* $\operatorname{acl}(A)$ *such that C and C_t have finite symmetric difference, the family $\mathcal{C}$ is generically very ample.*

Proof. Throughout, we assume $A = \varnothing$. It is immediate from Proposition 3.19 that (2) implies (1) and (1) implies (3). To see that (3) implies (1), simply note that every plane curve in X is a generic member of some $\varnothing$-definable family of plane curves, and the parameter space for such a family can be made stationary after passing to $\operatorname{acl}(\varnothing)$. So it is enough to show that (1) implies (2).

Now suppose (1) holds. As above, we can realize C as a generic member of a $\varnothing$-definable family of plane curves, say $C_{t_0} \in \mathcal{C} = \{C_t : t \in T\}$. By compactness and the definability of dimension, we may assume that for each $t \in T$, for any distinct generics $x, y \in C_t$, we have $\dim(t/xy) \leq \dim T - 2$ – that is, that $\{C_t : t \in T\}$ is generically very ample. By Proposition 3.19, $\mathcal{C}$ becomes very ample after removing a nongeneric $\varnothing$-definable subset of its graph. More precisely, apply Proposition 3.19 to the restriction of the family T to an $\operatorname{acl}(\varnothing)$-definable stationary generic subset $T_0 \subseteq T$ containing t_0, then use the fact that the conjugates of T_0 over $\varnothing$ are also very ample, and therefore, so is their union, T. Since $t_0 \in T$ is generic, only finitely many points of C_{t_0} are removed, which is enough to prove (2). □

We conclude this section with a geometric characterization of very ampleness.

Corollary 3.21. *Let X be strongly minimal. Then X is very ample if and only if there is a very ample definable family of plane curves in X.*

Proof. If X is very ample, by definition there is a plane curve in X which is very ample over some set. Then, by the previous corollary, there is a very ample family of plane curves in X.

If on the other hand there is a very ample family, then by the proposition there is a very ample plane curve over some set. Then by Corollary 3.14 every strongly minimal component of that plane curve is very ample over the same set, so that by definition X is very ample. $\square$

3E. *Preservation properties.* In this section, we develop three basic preservation properties of very ampleness; we then pose a fourth (stronger) property as a question, and answer this question in the context of expansions of fields. To begin we point out, generalizing the preservation under naming parameters (Lemma 3.11), that very ampleness is preserved under arbitrary expansions.

Lemma 3.22. *Any strongly minimal expansion of a very ample strongly minimal structure is very ample.*

Proof. Immediate by Corollary 3.21, since a very ample family is still definable in any expansion (technically we use here that the dimension of a definable set is unchanged in strongly minimal expansions, which is easy to check). $\square$

Next we show that very ample plane curves are closed under "independent compositions," in an appropriate sense. This will be used later on to show that very ample strongly minimal structures admit very ample families of plane curves of arbitrarily large dimension.

Lemma 3.23. *Let X be strongly minimal and A-definable. Let C and D be plane curves in X, definable over As and At, respectively. Assume that s and t are independent over A. If C and D are each very ample over A, then so is $D \circ C$.*

Proof. Without loss of generality, assume $A = \varnothing$. Let (x_1, z_1), (x_2, z_2) be distinct generics in $D \circ C$ over st. It will suffice to show that

$$\dim(st/x_1z_1x_2z_2) \le \dim(st) - 2.$$

Without loss of generality, assume $x_1 \ne x_2$. By definition of $D \circ C$ there are y_1, y_2 such that each $(x_i, y_i) \in C$ and each $(y_i, z_i) \in D$. Then each (x_i, y_i) is generic in C over st (so also over s), which by the very ampleness of C gives that

$$\dim(s/x_1y_1x_2y_2) \le \dim(s) - 2.$$

Next, note that y_i and z_i are interalgebraic over t (for $i = 1, 2$), by the nontriviality of D; see Remark 3.9. Thus,

$$\begin{aligned}\dim(st/x_1z_1x_2z_2) &= \dim(sty_1y_2/x_1z_1x_2z_2)\\ &= \dim(t/x_1z_1x_2z_2) + \dim(y_1y_2/tx_1z_1x_2z_2) + \dim(s/tx_1y_1z_1x_2y_2z_2).\end{aligned}$$

In this last expression, note that the first term is at most $\dim(t)$, the second term is 0, and the third term is at most $\dim(s)-2$ as we have seen above. Thus,

$$\dim(st/x_1z_1x_2z_2) \leq \dim(s)+\dim(t)-2,$$

which, by the independence of s and t, is equivalent to the desired statement. $\square$

Finally, we point out that very ampleness is preserved under finite-to-one functions.

Lemma 3.24. *Suppose X and Y are strongly minimal sets, and $f : X \to Y$ is a definable finite-to-one function. If X is very ample, then so is Y.*

Proof. Without loss of generality, assume X, Y, and f are $\varnothing$-definable. Let C be a strongly minimal plane curve in X which is very ample over some set A. Without loss of generality, we assume also that $A=\varnothing$. Let $c=\mathrm{Cb}(C)$. After editing finitely many points, we may assume C is definable over c.

Now let D be the image of C in Y^2 (applying f to each coordinate). So D is definable over c. Since f is finite-to-one, D is moreover strongly minimal. We claim that D is very ample in Y. To that end, suppose z, w are distinct generics in D over c. It will suffice to show that z forks with c over w. Now by assumption there are $x, y \in C$ with $f(x)=z$ and $f(y)=w$. Since f is finite-to-one, each of the pairs x, z and y, w are interalgebraic. So by dimension considerations, x and y are generics in C. Moreover, since $z \neq w$, it follows that $x \neq y$. Thus, by the very ampleness of C, x forks with c over y. Then, by interalgebraicity, z forks with c over w, as desired. $\square$

Recall that two strongly minimal sets are *nonorthogonal* if there is a definable (possibly over additional parameters) finite-to-finite correspondence between them. Lemma 3.24 does not extend to the preservation of very ampleness under nonorthogonality, as can be seen from Example 3.6. Indeed, the (non-very ample) structure $\mathcal{M}$ from Example 3.6 is nonorthogonal to the strongly minimal set $K/{\sim}$, where $\sim$ is the relation $x^2=y^2$ on K; but the induced structure on $K/{\sim}$ is a pure algebraically closed field, so very ample by Example 3.5.

As we will see later, the main reason very ampleness does not pass from $K/{\sim}$ to $\mathcal{M}$ in Example 3.6 is that K is not *internal* to $K/{\sim}$ (in the sense of the reduct, $\mathcal{M}$). We may thus ask the following:

Question 3.25. Is very ampleness preserved under internality? That is, if X is very ample, and Y is strongly minimal and internal to X, must Y be very ample?

To conclude this subsection, we now answer Question 3.25 whenever X has a definable field structure. We will later improve on this result in Corollary 4.2 and in the proof of Proposition 6.3. In order to smoothly recall the setting in this later proposition, we present the result here as a corollary of two general statements, Lemma 3.27 and Proposition 3.28.

Before proceeding, let us clarify:

Remark 3.26. By *expansion* below, we are assuming the field structure is part of the signature of $(K, +, \cdot, \dots)$. In general, for a strongly minimal structure, $\mathcal{M}$, admitting a definable field structure (with universe M), elimination of imaginaries will hold after naming any set of parameters defining a field structure on M.

Lemma 3.27. *Every strongly minimal expansion of an algebraically closed field eliminates imaginaries.*

Proof. Let $(K, +, \cdot, \dots)$ be an expansion of an algebraically closed field. Because of the field structure, it is automatic that $\operatorname{acl}(\varnothing)$ is infinite. So by weak elimination of imaginaries in strongly minimal structures, it suffices to code finite sets: that is, any such expansion $(K, +, \cdot, \dots)$ has elimination of imaginaries if and only if there are $\varnothing$-definable injections $K^{(n)} \to K^m$ for all n (and some m depending on n), where $K^{(n)}$ denotes the n-th symmetric power of K. But such functions already exist in the pure field $(K, +, \cdot)$, so we are done. $\square$

Proposition 3.28. *Let* $(K, +, \cdot, \dots)$ *be a strongly minimal expansion of an algebraically closed field, and let* $X \subset K^n$ *be a* $\varnothing$*-definable set of dimension* $r \geq 1$. *Let* $\mathcal{H} = \{H_t : t \in T\}$ *be the family of hyperplanes in* K^n, *and let* $t \in T$ *be generic. Then* $\dim(H_t \cap X) = r - 1$, *and if* $x, y \in H_t \cap X$ *are distinct generic over* t, *then* x *and* t *fork over* y.

Proof. Throughout this proof, we freely use the fact that the strongly minimal expansion $(K, +, \cdot, \dots)$ preserves the dimensions of $(K, +, \cdot)$-definable sets, which is well known and easy to check. We also use that T can be identified by a large subset of $\mathbb{P}^n(K)$ (the set of projective hyperplanes intersecting K^n); in particular, it follows that T is stationary of dimension n. As is well known, for any $x \in K^n$ the set H^x (the hyperplanes through x) then has dimension $n-1$, and for $x \neq y$ the set $H^x \cap H^y$ has dimension $n-2$.

Now we prove the proposition in two claims.

Claim 3.28.1. $$\dim(H_t \cap X) = r - 1.$$

Proof. Let $x \in X$ be generic and $s \in H^x$ be generic over x. Then $\dim(xs) = r + n - 1$. Since $\dim(s) \leq n$, this forces $\dim(x/s) \geq r - 1$. In particular, $\dim(H_s \cap X) \geq r - 1$.

Let us verify that $\dim(H_s \cap X) = r - 1$. Suppose not. Then there is $y \in H_s \cap X$ with $\dim(y/xs) = r$. Thus $\dim(xys) = 2r + n - 1$. But $\dim(xy) \leq 2r$, so $\dim(s/xy) \geq n - 1$. As observed above, this is only possible if $x = y$; but clearly $x \neq y$, since $\dim(y/xs) = r \geq 1$.

So $\dim(H_s \cap X) = r - 1$, and thus $\dim(x/s) \leq r - 1$. Recalling that $\dim(xs) = r + n - 1$, we conclude that $\dim(s) = n$ and $\dim(x/s) = r - 1$. In particular, s is generic in T. Since T is stationary, $\operatorname{tp}(s) = \operatorname{tp}(t)$, so also $\dim(H_t \cap X) = r - 1$. $\square$

Claim 3.28.2. *Let* $x \neq y$ *be generics in* $H_t \cap X$ *over* t. *Then* x *forks with* t *over* y.

Proof. By assumption $\dim(t) = n$, and by the previous claim, $\dim(y/t) = r - 1$. So $\dim(yt) = n + r - 1$. But $\dim(t) \leq r$ since $y \in X$; and since $t \in H^y$, we get $\dim(t/y) \leq n - 1$. So we must have $\dim(y) = r$ and $\dim(t/y) = n - 1$. But since $t \in H^x \cap H^x$, we also have $\dim(t/xy) \leq n - 2$, which proves the claim. □

Thus we have proved Proposition 3.28. □

Corollary 3.29. *Let K be a strongly minimal expansion of a field, and let X be a strongly minimal set internal to K. Then X is very ample.*

Proof. Adding parameters if necessary, we may assume X is $\varnothing$-definable. By Lemma 3.27, we may assume $X \subset K^n$ for some n. Then by Proposition 3.28 (applied to $X^2 \subset K^{2n}$), there is a very ample plane curve in X. So X is very ample. □

3F. ***When very ampleness can be guaranteed.*** As we have seen, our first example of a non-very ample structure (Example 3.6) still admits a very ample *sort* — that is, it becomes very ample after dividing by a definable equivalence relation. We thus think of K of that example as very ample "up to a finite cover," which is precisely the situation of Theorem B in [Hrushovski and Zilber 1996]. We do not know whether this is true of all strongly minimal nonlocally modular structures.

Question 3.30. Suppose X is strongly minimal and not locally modular.

(1) Can one always find a very ample strongly minimal set Y internal to X?

(2) More specifically, can one always find a definable equivalence relation $\sim$ on X so that $X/\sim$ is strongly minimal and very ample?

Later in this paper, we will see that for strongly minimal expansions of groups G, the answer to Question 3.30 (2) is positive. To show that, we pick a well-chosen two-dimensional family of plane curves $\{C_t : t \in T\}$ and show that the relation $I(x, y) \subset G^2$, given by $I(x, y)$ if and only if for some z and w we have $|C^{(x,z)} \cap C^{(y,w)}| = \infty$, is contained in the graph of an equivalence relation $\sim$ on G with finite classes, (in fact, the equivalence classes are cosets of a finite subgroup). Then we show that in the quotient $G/\sim$ the image of $\{C_t : t \in T\}$ is very ample. For a general strongly minimal structure, this strategy seems to work once we can find a definable two-dimensional family of plane curves where the analogous relation I is contained in an equivalence relation with finite classes. It can also be shown that we can always find families of plane curves where $I(x, y)$ is one-dimensional. However, we do not know how to extend I to an equivalence relation in general.

Proposition 3.32 shows that, as one might expect from the above discussion, nonlocally modular strongly minimal structures with elimination of imaginaries are very ample.

Definition 3.31. Let X be strongly minimal and A-definable. We say that X *eliminates imaginaries over* A if for any elements $a_1, \dots, a_n \in X$ and any $c \in \mathrm{dcl}^{eq}(A\bar{a})$, there is a finite sequence $b_1, \dots, b_m \in X$ with $\mathrm{dcl}^{eq}(A\bar{b}) = \mathrm{dcl}^{eq}(Ac)$.

Proposition 3.32. *Let X be strongly minimal, A-definable, and not locally modular. If X eliminates imaginaries over A, then X is very ample.*

Proof. By nonlocal modularity, there are a set $B \supset A$ and a strongly minimal plane curve $C \subset X^2$, such that $\dim(c/B) = 2$, where $c = \mathrm{Cb}(C)$. By stable embeddedness of X, there are $a_1, \dots, a_n \in X$ with $c \in \mathrm{dcl}(A\bar{a})$. So by elimination of imaginaries, there are $b_1, \dots, b_m \in X$ with $\mathrm{dcl}(A\bar{b}) = \mathrm{dcl}(Ac)$, and in particular, $\mathrm{dcl}(B\bar{b}) = \mathrm{dcl}(Bc)$. So $\dim(\bar{b}/B) = 2$, and we may extract a two element basis for $\bar{b}$ over B. Without loss of generality, let us assume $\dim(b_1b_2/B) = 2$. So $b_1b_2 \in \mathrm{dcl}(Bc)$ and $c \in \mathrm{acl}(Bb_1b_2)$.

Now let $x \in X^2$ be generic in C over Bc. So $\dim(cx/B) = 3$. Since $c \notin \mathrm{acl}(B)$, x forks with c over B, which implies that $\dim(x/B) = 2$, and thus $\dim(c/Bx) = 1$. By interalgebraicity, $\dim(b_1b_2/Bx) = 1$. Let $p = \mathrm{stp}(b_1b_2/Bx)$; so p is a minimal type in X^2. Let S be a strongly minimal plane curve whose generic type is p. So (b_1, b_2) is generic in S over $\mathrm{acl}(Bx)$.

Claim 3.32.1. *S is very ample in X over B.*

Proof. Let (b_1', b_2') be another generic of S over $\mathrm{acl}(Bx)$. So $(b_1', b_2') \models p$. Then there is some c' with $\mathrm{tp}(b_1b_2cx/B) = \mathrm{tp}(b_1'b_2'c'x/B)$. Since $b_1b_2 \in \mathrm{dcl}(Bc)$ and $(b_1', b_2') \neq (b_1, b_2)$, it follows that $c \neq c'$. Because $\mathrm{tp}(cx) = \mathrm{tp}(c'x)$ there is an automorphism mapping cx to $c'x$ and mapping $\mathrm{tp}(x/c)$ to $\mathrm{tp}(x/c')$. Since $c \neq c'$, $c = \mathrm{Cb}(x/c)$, and $c' = \mathrm{Cb}(x/c')$, the (stationary) types $\mathrm{tp}(x/c)$ and $\mathrm{tp}(x/c')$ have no common nonforking extension (essentially, by the definition of canonical bases), implying that x forks with c' over c. □

We have shown that X admits a very ample strongly minimal plane curve, so X is very ample. □

Remark 3.33. The formulation of the above proof using types is convenient, but may obscure the geometric idea. Let us now explain the proof in more geometric terms: we first use nonlocal modularity to find a faithful two-dimensional family of plane curves; we then use elimination of imaginaries to reparametrize this family by a definable set $T \subset X^n$, and by taking a projection $X^n \to X^2$, we further reparametrize (up to losing generic strong minimality of the curves) by X^2. We then "dualize" the resulting family (interchanging points and curves), and observe that the "dual" notion of faithfulness is precisely very ampleness.

Remark 3.34. The converse of Proposition 3.32 is false: Let $\mathcal{M} = (M, \dots)$ be the full $\mathbb{C}$-induced structure on a smooth irreducible curve M over $\mathbb{C}$ of genus $g > 0$. Then $\mathcal{M}$ is very ample by Corollary 3.29, but does not eliminate the imaginary $\mathbb{C}$

over any set of parameters. Let us elaborate: First, by the Zilber trichotomy ([Hasson and Sustretov 2017] or [Castle 2023]), $\mathcal{M}$ interprets a set isomorphic to $\mathbb{C}$, and thus interprets a smooth curve X of genus 0. If X were embeddable into some M^n, then by composing with a projection one could obtain a nonconstant rational map $X \to M$, which (by completing, normalizing, and then comparing the genus of each curve) is easily seen to contradict the Riemann–Hurwitz formula.

4. Applications of very ampleness

We now turn toward applications of very ampleness, with an eye toward improved versions of the Zilber trichotomy for structures interpreted in algebraically closed fields. Unless explicitly stated otherwise, we work in a uncountable saturated stable structure $\mathcal{M}$. Throughout this section and Section 7 statements will often include the disclaimer "assuming Zilber's trichotomy" (for structures interpretable in algebraically closed fields) or the weaker "an algebraically closed field is interpretable". At the time of writing of this paper, the status of Zilber's trichotomy for such structures is as follows:

(1) Hasson and Sustretov [2017] have a proof of the trichotomy in the one-dimensional case.

(2) Castle [2023] has a proof of the trichotomy in characteristic 0 (for all dimensions). This result is independent of [Hasson and Sustretov 2017].

(3) These results are yet to be published.

(4) Castle, Hasson, and Ye [Castle et al. 2024] have a proof of the trichotomy in positive characteristic.

4A. ***Very ampleness and internality.*** The following proposition summarizes the effect of very ampleness in comparing strongly minimal sets. In fact, almost every application we make of very ampleness will, ultimately, follow from this result:

Proposition 4.1. *Let X and Y be strongly minimal sets, and assume X is very ample. If X is nonorthogonal to Y, then X is internal to Y.*

Proof. By nonorthogonality, there is a strongly minimal set $C \subset X \times Y$ such that both projections $C \to X$ and $C \to Y$ are finite-to-one. Without loss of generality (by adding finitely many points to C), assume that $C \to X$ is surjective. We then view C as a multivalued function $X \to Y$, mapping each $x \in X$ to the finite set $\{y \in Y : (x, y) \in C\}$. In particular, it follows that one can find a sort S in Y^{eq}, and a finite-to-one definable map $f : X \to S$. Absorbing parameters into the language, we may assume that f is $\varnothing$-definable.

Now by assumption there is a plane curve $C \subseteq X^2$ which is very ample over some set A. Applying f coordinatewise, we obtain a finite-to-one function $\tilde{f} : C \to S^2$.

Note that since f is $\varnothing$-definable, if $\tilde{f}(x_1, x_2) = \tilde{f}(y_1, y_2)$ then $(x_1, x_2), (y_1, y_2)$ are interalgebraic over $\varnothing$. So by condition (3) of Lemma 3.11, there are no distinct generics $(x_1, x_2), (y_1, y_2) \in C$ over A such that $\tilde{f}(x_1, x_2) = \tilde{f}(y_1, y_2)$. In other words, $\tilde{f}$ is generically injective. Letting $T = \mathrm{im}(\tilde{f})$ (potentially minus a finite set) and inverting $\tilde{f}$, we can then extract a definable function $g : T \to C$ with cofinite image. Then composing with a coordinate projection π, the resulting map $\pi \circ g : T \to X$ has cofinite image in X. This shows that X is internal to T, and since $T \subset S^2$, T is internal to Y. Thus, X is internal to Y. □

Before moving on we give two quick applications. First, the following is immediate from Proposition 4.1, Lemma 3.27, and Corollary 3.29:

Corollary 4.2. *Let K be a strongly minimal expansion of an algebraically closed field, and let X be a strongly minimal set nonorthogonal to K. Then the following are equivalent:*

(1) *X is very ample.*

(2) *X is internal to K.*

(3) *There is a definable embedding of X into some K^n.*

It follows from this last corollary, that if X is very ample and interprets an algebraically closed field, K, then any strongly minimal set Y interpretable in X is very ample. Indeed, by the corollary X is internal to K, and since Y is internal to X, Y too is internal to K, so by the corollary, again, Y is very ample. Thus, in particular, a counterexample to Question 3.25 must be a counterexample to Zilber's trichotomy.

Next, we give a partial answer to Question 3.30 for strongly minimal sets nonorthogonal to very ample sets.

Corollary 4.3. *Let X and Y be strongly minimal and nonorthogonal, and assume Y is very ample. Then:*

(1) *There is a very ample strongly minimal set internal to X.*

(2) *If, moreover, Y eliminates imaginaries (over some parameter set), then there is a definable equivalence relation $\sim$ on X with finite classes such that $X/\sim$ is very ample.*

Proof. (1) is immediate, because by Proposition 4.1, Y is internal to X. So we need only prove (2). For that, assume that Y eliminates imaginaries over some set. By nonorthogonality, one can find a finite-to-one map $f : X \to S$, where S is a sort in Y^{eq}. By elimination of imaginaries we may assume $S \subset Y^n$ for some n, so we have $f : X \to Y^n$. Then composing with an appropriate projection $Y^n \to Y$, we obtain a finite-to-one map $X \to Y$, which shows that (up to editing finitely many points) Y is a quotient of X. □

Remark 4.4. In particular, it follows from Corollary 4.3 that every nonlocally modular strongly minimal structure satisfying the Zilber trichotomy admits a positive answer to both parts of Question 3.30.

4B. ***Very ample Zariski geometries.*** In the present section, we show that our definition of very ampleness is equivalent, in the context of Zariski geometries, to the original definition of Hrushovski and Zilber. The results of this section are not needed in the sequel, and the reader only interested in the application may safely skip to the next section.

Recall that a strongly minimal Zariski geometry[4] is a structure $\mathcal{Z}$, in a language equipping it with a compatible system of Noetherian topologies on Z^n (all n), such that $\mathcal{Z}$ has quantifier elimination and satisfies the dimension theorem; see [Hrushovski and Zilber 1996] for the details. The Zariski geometry $\mathcal{Z}$ is ample if it is not locally modular; it is then shown [Hrushovski and Zilber 1996, Theorem B] that any ample Zariski geometry interprets an algebraically closed field K. To further identify $\mathcal{Z}$ with a smooth curve over K, the notion of very ampleness is introduced. We now give the definition (paraphrased) as it appears in [Hrushovski and Zilber 1996]. For the sake of clarity, we will refer to this version as Z-very ample.

Definition 4.5. Let $\mathcal{Z}$ be a Zariski geometry with universe Z. Then $\mathcal{Z}$ is *Z-very ample* if there are an irreducible closed set $E \subset Z^n$ for some n, and an irreducible closed set $C \subset E \times Z^2$, such that the following hold:

(1) For generic $e \in E$, the set $C(e)$ is irreducible and one-dimensional.

(2) For any distinct $a, b \in Z^2$ there is some $e \in E$ such that $C(e)$ contains exactly one of a, b.

The main theorem in [Hrushovski and Zilber 1996], equivalently stated, then asserts that every Z-very ample Zariski geometry is isomorphic (as a Zariski geometry) to a smooth curve over an algebraically closed field.

It is not directly mentioned in [Hrushovski and Zilber 1996] that the converse holds — namely, that every smooth curve is very ample. We thank Hrushovski for outlining the following proof:

Lemma 4.6. *Let Z be a smooth algebraic curve over an algebraically closed field. Then Z is Z-very ample.*

Proof. Since C is a curve it is quasiprojective; so we may assume $C \subset \mathbb{P}^n(K)$ for some n. Our goal is to find a family $C \subset E \times Z^2$ consisting of all hyperplane intersections in Z^2 (that is, sets $H_t \cap Z^2$ where H_t is a hyperplane in $\mathbb{P}^n(K)$). In this case, (1) follows by Bertini's theorem, and (2) follows since any two points in projective space can be separated by a hyperplane.

[4]We are referring to the term as appearing in [Hrushovski and Zilber 1996] and not to the more general one of Zilber [2010].

The only difficulty here is finding such a family which is irreducible, closed, and parametrized by a closed irreducible set in a power of Z (rather than the natural parametrization by $\mathbb{P}^n(K)$). For this, it will suffice to find a surjective morphism $Z^n \to \mathbb{P}^n(K)$, since then we can lift the natural parametrization from $\mathbb{P}^n(K)$ to Z^n.

To find such a morphism, we note that by composing with a surjection $(\mathbb{P}^1)^n \to \mathbb{P}^n$ (which exists because the former is a projective variety of dimension n), it will moreover suffice to find a surjection $C \to \mathbb{P}^1(K)$.

Finally, let us build a surjection $C \to \mathbb{P}^1(K)$. First choose any (dominant) morphism $f : C \to \mathbb{P}^1(K)$, and suppose $f(C)$ misses m points of $\mathbb{P}^1(K)$, say $x_1, \ldots, x_m$. Then the map $g \circ f : C \to \mathbb{P}^1(K)$ is surjective, where $g : \mathbb{P}^1(K) \to \mathbb{P}^1(K)$ is any ramified cover with generically $(m+1)$-sized fibers that does not ramify at any of $x_1, \ldots, x_m$. □

It follows by [Hrushovski and Zilber 1996] and Lemma 4.6 that a Zariski geometry $\mathcal{Z}$ is Z-very ample if and only if it is isomorphic to a smooth curve over an algebraically closed field. Our goal now is to show that the same statement holds using our notion of very ampleness in place of Z-very ampleness. To do this, we first need to inspect the proof in [Hrushovski and Zilber 1996], in order to extract the precise use of Z-very ampleness and replace it with a use of very ampleness. In fact, the proof in [Hrushovski and Zilber 1996] is achieved by separately proving the following three facts:

Fact 4.7. Let $\mathcal{Z}$ be a Zariski geometry with universe Z.

(1) If $\mathcal{Z}$ is not locally modular, then $\mathcal{Z}$ interprets an algebraically closed field (p. 47, first paragraph of the proof of Theorem A).

(2) If $\mathcal{Z}$ interprets the algebraically closed field K, and $\mathcal{Z}$ is Z-very ample, then Z is internal to K (from (1) to the end of the paragraph splitting p. 47–48).

(3) If $\mathcal{Z}$ interprets the algebraically closed field K, and $\mathcal{Z}$ is internal to K, then $\mathcal{Z}$ is isomorphic to a smooth curve over K (p. 48, the rest of the proof of Theorem A).

We now show:

Proposition 4.8. *Let $\mathcal{Z}$ be a Zariski geometry. Then the following are equivalent*:

(1) *$\mathcal{Z}$ is very ample.*

(2) *$\mathcal{Z}$ is Z-very ample.*

(3) *$\mathcal{Z}$ is isomorphic to a smooth curve over an algebraically closed field.*

Proof. As stated above, the equivalence of (2) and (3) is [Hrushovski and Zilber 1996, Theorem A] and Lemma 4.6. Moreover, the implication (3) $\Longrightarrow$ (1) follows from Corollary 3.29. So it will suffice to show (1) $\Longrightarrow$ (3).

Assume (1). Then by Lemma 3.8, $\mathcal{Z}$ is not locally modular. So by Fact 4.7 (1), $\mathcal{Z}$ interprets an algebraically closed field K. Then by Proposition 4.1, Z is internal to K. So by Fact 4.7 (3), $\mathcal{Z}$ is isomorphic to a smooth curve over K. Thus (3) holds, and we are done. □

4C. *Very ample strongly minimal ACF-relics.* We now turn toward applications of very ampleness to structures interpretable in algebraically closed fields. In [Loveys 2004, Theorem 7.1] Loveys gives a complete list (up to definable finite covers) of all locally modular nontrivial reducts of the algebraically closed field $\mathbb{C}$. If we are interested in the classification of nonlocally modular reducts of $\mathbb{C}$ only up to finite covers, then Rabinovich's theorem [1993] is the final word. But one could hope for a more precise classification of such reducts. One natural question is the identification of reducts that are not proper. To state this problem in somewhat greater generality, it is convenient to have:

Definition 4.9. Let $\mathcal{N}$ be any structure.

(1) An $\mathcal{N}$-*relic* is a reduct, $\mathcal{M}$, of the structure induced on some $\mathcal{N}$-definable set M.

(2) If $\mathcal{M}$ is an $\mathcal{N}$-relic, and X is $\mathcal{M}$-definable, then X is $\mathcal{N}/\mathcal{M}$-*full* if every $\mathcal{N}$-definable subset of every X^n is $\mathcal{M}$-definable.

(3) If $\mathcal{M} = (M, \dots)$ is an $\mathcal{N}$-relic, then $\mathcal{M}$ is *full in* $\mathcal{M}$ if the universe M is $\mathcal{N}/\mathcal{M}$-full.

In a recent unpublished work, Castle and M. Tran give a concrete characterization of fullness for reducts of $\mathbb{C}$ whose atomic sets are polynomial functions, though a similarly concrete identification of all full reducts of $\mathbb{C}$ is not available. In the next two subsections, we show, roughly, that the fullness of an ACF-relic is equivalent to very ampleness. In fact, in the main results of the current subsection, we show an equivalence between several properties of a strongly minimal ACF-relic, notably very ampleness, fullness, and bi-interpretability with the field.

We begin with a couple of straightforward and well-known facts about relics.

Lemma 4.10. *Let $\mathcal{M}$ be an $\mathcal{N}$-relic, and let X be $\mathcal{M}$-definable and $\mathcal{N}/\mathcal{M}$-full. If Y is $\mathcal{M}$-definable and internal to X* (*in the sense of $\mathcal{M}$*), *then Y is also $\mathcal{N}/\mathcal{M}$-full.*

Proof. Let $Z \subset Y^n$ be $\mathcal{N}$-definable. By internality, there are an $\mathcal{M}$-definable set $W \subset X^m$, for some m, and an $\mathcal{M}$-definable surjective function $f : W \to Y^n$. By the fullness of X, $f^{-1}(Z) \subset X^m$ is $\mathcal{M}$-definable. Then since f is surjective, $Z = f(f^{-1}(Z))$ is also $\mathcal{M}$-definable. Thus Y is $\mathcal{N}/\mathcal{M}$-full. □

Lemma 4.11. *Let $(K, +, \cdot)$ be an algebraically closed field, let $\mathcal{M} = (M, \dots)$ be a K-relic, and F an infinite $\mathcal{M}$-definable field. Then F is a pure field as seen from both K and $\mathcal{M}$. In particular, F is $K/\mathcal{M}$-full.*

Proof. Since $\mathcal{M}$ defines fewer sets than K, and $\mathcal{M}$ defines the field operations on F, it suffices to show that every K-definable subset of every F^n is definable from only the field operations on F. So let $X \subset F^n$ be K-definable. By [Poizat 1988], there is a K-definable field isomorphism $f : K \to F$. Then $f^{-1}(X) \subset K^n$ is definable from the field operations on K, and since f is an isomorphism, X is definable from the field operations on F. □

Using the previous two lemmas, we now give several equivalent characterizations of fullness for ACF-relics. These are likely well known.

Proposition 4.12. *Let K be an algebraically closed field, and let $\mathcal{M} = (M, \dots)$ be a K-relic. Then the following are equivalent*:

(1) *$\mathcal{M}$ is full in K.*

(2) *$\mathcal{M}$ is bi-interpretable with K over parameters.*

(3) *There is an infinite field F definable in $\mathcal{M}$, such that M is internal to F (in the sense of $\mathcal{M}$).*

(4) *$\mathcal{M}$ defines an infinite field, and for every infinite $\mathcal{M}$-definable field F, M is internal to F (in the sense of $\mathcal{M}$).*

Proof. (1) $\Longrightarrow$ (2): Assume $\mathcal{M}$ is full. First, we point out that $\mathcal{M}$ interprets an infinite field. Indeed, let X be any strongly minimal set in $\mathcal{M}$. By fullness, X is also strongly minimal in K. Thus, by quantifier elimination in K, after potentially deleting finitely many points we may assume X is a smooth irreducible curve, and we conclude by the main theorem on Zariski geometries [Hrushovski and Zilber 1996].

Now let F be an infinite $\mathcal{M}$-definable field. By [Poizat 1988], there is a K-definable isomorphism $f : K \to F$. Let $\mathcal{Y} = (Y, \dots)$ be the image of $\mathcal{M}$ under f (equivalently, the relic of F defined using the same formulas defining $\mathcal{M}$ in K).

We claim that the given interpretations of $\mathcal{M}$ in K, and of F in $\mathcal{M}$, provide a bi-interpretation of K and $\mathcal{M}$ over parameters. Indeed, we already know that K and F are definably isomorphic in K, as witnessed by f. So it remains to show that $\mathcal{Y}$ is $\mathcal{M}$-definably isomorphic to $\mathcal{M}$. But f gives a K-definable isomorphism $\mathcal{M} \to \mathcal{Y}$, and $\mathcal{Y}$ resides entirely in $\mathcal{M}^{\mathrm{eq}}$. So by the fullness of $\mathcal{M}$, this isomorphism is $\mathcal{M}$-definable.

(2) $\Longrightarrow$ (3): The statement of (3) is part of the definition of a bi-interpretation.

(3) $\Longrightarrow$ (4): Assuming (3), let F_1 be an infinite $\mathcal{M}$-definable field such that M is internal to F_1 in the sense of $\mathcal{M}$. To prove (4), it remains to show that if $\mathcal{M}$ defines another infinite field, say F_2, then M is internal to F_2.

By assumption, F_2 is internal to F_1, and thus nonorthogonal to F_1. But by Lemma 4.11, each F_i is very ample and strongly minimal. So by Proposition 4.1, F_1 is internal to F_2, and thus so is M.

(4) $\implies$ (1): Assume (4). Let F be an infinite $\mathcal{M}$-definable field. By assumption, M is internal to F in the sense of $\mathcal{M}$. By Lemma 4.11, F is $K/\mathcal{M}$-full. So by Lemma 4.10, M is $K/\mathcal{M}$-full, i.e., $\mathcal{M}$ is full in K. □

Remark 4.13. Given a K-relic $\mathcal{M}$, fullness of $\mathcal{M}$ may, a priori. depend on the interpretation of $\mathcal{M}$ in K. However, as the other five conditions appearing in the statement do not depend on the interpretation, it follows that fullness is a property only of the abstract structure $\mathcal{M}$. In other words, if $\mathcal{M}$ is interpretable in an algebraically closed field, then either all possible interpretations are full or none are.

We now use Proposition 4.12 and the Zilber trichotomy to show that very ampleness characterizes fullness for strongly minimal relics. This can be seen as an analogue of the main theorem on Zariski geometries.

Theorem 4.14. *Let K be an algebraically closed field, and let $\mathcal{M} = (M, \dots)$ be a strongly minimal K-relic which satisfies Zilber's trichotomy. Then the following are equivalent*:

(1) *$\mathcal{M}$ is very ample.*

(2) *$\mathcal{M}$ is full in K.*

(3) *Up to deleting a finite set, $\mathcal{M}$ is isomorphic to an irreducible algebraic curve over K, with its induced structure from K.*

Proof. (1) $\implies$ (2): Assume $\mathcal{M}$ is very ample. By Lemma 3.8, $\mathcal{M}$ is not locally modular, so by the Zilber trichotomy, $\mathcal{M}$ interprets an infinite field, say F. Then M and F are nonorthogonal, and by Lemma 4.11, F is strongly minimal. So by Proposition 4.1, M is internal to F. Then, by Proposition 4.12, $\mathcal{M}$ is full.

(2) $\implies$ (3): Assuming $\mathcal{M}$ is full, the universe M is also strongly minimal in K. So by quantifier elimination in K, M agrees up to finitely many points with an irreducible curve. This makes (3) clear.

(3) $\implies$ (1): Follows by Corollary 3.29. □

4D. *More on full ACF-relics.* It follows from Theorem 4.14 that strongly minimal relics of algebraically closed fields are full precisely when they are very ample. In this section, we give a similar characterization of fullness for a general (nonstrongly minimal) relic. In this case, the argument will be a little more delicate. In particular, we need to first develop general conditions under which fullness of a relic can be checked at the level of strongly minimal sets.

Proposition 4.15. *Let $\mathcal{M}$ be an $\mathcal{N}$-relic, where $\mathcal{M}$ and $\mathcal{N}$ are almost strongly minimal. Assume there is a strongly minimal set in $\mathcal{M}$ which is $\mathcal{N}/\mathcal{M}$-full. Then the following are equivalent*:

(1) *$\mathcal{M}$ is full in $\mathcal{N}$.*

(2) *Every strongly minimal set in $\mathcal{M}$ is $\mathcal{N}/\mathcal{M}$-full.*

(3) *Every strongly minimal set in $\mathcal{M}$ is also strongly minimal in $\mathcal{N}$.*

(4) *Every stationary definable set in $\mathcal{M}$ is also stationary in $\mathcal{N}$.*

Proof. Throughout, we use the following:

Claim 4.15.1. *Let X be any $\mathcal{M}$-definable set. Then* $\dim_{\mathcal{M}}(X) = \dim_{\mathcal{N}}(X)$.

Proof. Let F be an $\mathcal{N}/\mathcal{M}$-full strongly minimal set. By almost strong minimality, there is an $\mathcal{M}$-definable set Y internal to F (in the sense of $\mathcal{M}$) which is in $\mathcal{M}$-definable finite correspondence with X. Since finite correspondences preserve dimension, we may assume $X = Y$, i.e., that X is internal to F. But then by Lemma 4.10, X is $\mathcal{N}/\mathcal{M}$-full, which makes the lemma obvious. ☐

Notation. In light of Claim 4.15.1, for the rest of the proof we will drop all subscripts when denoting dimensions of definable sets.

We now prove the proposition.

(1) $\Longrightarrow$ (2): Clear.

(2) $\Longrightarrow$ (3): Clear.

(3) $\Longrightarrow$ (4): Assume (3), and let X be stationary in the sense of $\mathcal{M}$. If $\dim(X) = 0$, everything is clear, so assume $\dim(X) = d \geq 1$. Using coordinatization in almost strongly minimal structures, one can $\mathcal{M}$-definably write $X = \bigcup_{t \in T} X_t$, where T is strongly minimal in $\mathcal{M}$, for any $t \in T$ we have $\dim(X_t) = d-1$, and X_t is stationary in $\mathcal{M}$ for generic $t \in T$. Specifically, after possibly adding a finite set of parameters, there is a strongly minimal set T_0 such that $M = \operatorname{acl}(T_0)$ so given a generic $x \in X$ there is $t_0 \in T_0$ such that x forks over t_0. Take $c := \operatorname{Cb}(\operatorname{stp}(x/t_0))$, T strongly minimal in $\operatorname{stp}(c)$, and $\varphi(x,t)$ a formula isolating $\operatorname{stp}(x/t_0)$ in its dimension and degree, X_t its set of realizations. Now, by (3), T is also strongly minimal in $\mathcal{N}$. Moreover, arguing by induction, we may assume that X_t is stationary in $\mathcal{N}$ for generic $t \in T$. Then, as the union of a stationary family of (generically) stationary sets, it follows easily that X is stationary in $\mathcal{N}$.

(4) $\Longrightarrow$ (1): Assume (4), and let $X \subset M^n$ be $\mathcal{M}$-definable. We argue by induction on $d = \dim(X)$. Let F be an $\mathcal{N}/\mathcal{M}$-full strongly minimal set. By almost strong minimality, there are a sort S in F^{eq} and a finite-to-one $\mathcal{M}$-definable function $f : M^n \to S$. Let $Y = f^{-1}(f(X))$. Note that $f(X)$ is $\mathcal{M}$-definable, by the fullness of F; thus Y is also $\mathcal{M}$-definable. Moreover, $X \subset Y$, and since f is finite-to-one, we have $\dim(X) = \dim(Y) = d$.

In the structure $\mathcal{M}$, write $Y = Y_1 \cup \cdots \cup Y_m$ as a union of stationary sets of dimension d. By (4), each Y_i is stationary in $\mathcal{N}$, so each Y_i is either almost contained in X or almost disjoint from X. Since $X \subset Y$ and $\dim(X) = \dim(Y)$, X is almost equal to the union (say Z) of those Y_i that are almost contained in X. But Z is

$\mathcal{M}$-definable, and X is a Boolean combination of Z, $X - Z$, and $Z - X$, the latter two having dimension less than d. It then follows by induction that X is $\mathcal{M}$-definable. $\square$

We are now ready for the main theorem of this subsection.

Theorem 4.16. *Let K be an algebraically closed field, and assume the Zilber trichotomy holds for strongly minimal K-relics. Let $\mathcal{M} = (M, \dots)$ be a K-relic. Then the following are equivalent*:

(1) *$\mathcal{M}$ is full.*

(2) *There is a very ample strongly minimal set X in $\mathcal{M}$ such that M is internal to X (in the sense of $\mathcal{M}$).*

(3) *$\mathcal{M}$ is almost strongly minimal, and every strongly minimal set in $\mathcal{M}$ is very ample.*

Proof. (1) $\Longrightarrow$ (2): By Proposition 4.12, M is internal to an $\mathcal{M}$-definable infinite field, say F. By Lemma 4.11, F is strongly minimal and very ample.

(2) $\Longrightarrow$ (3): Let X be as in (2). Internality to X is a strengthening of almost strong minimality, so to show (3), it suffices to show that strongly minimal sets in $\mathcal{M}$ are very ample. So, let Y be strongly minimal in $\mathcal{M}$. By assumption, Y is internal to X in the sense of $\mathcal{M}$. So by Lemma 4.10, Y is $K/\mathcal{M}$-full. Then by Theorem 4.14 (applied to Y with its induced structure from $\mathcal{M}$), Y is very ample.

(3) $\Longrightarrow$ (1): Assume (3). We show fullness using Proposition 4.15. Since $\mathcal{M}$ and K are almost strongly minimal by assumption, it suffices to show that every strongly minimal set in $\mathcal{M}$ is $K/\mathcal{M}$-full. But this is automatic: if X is strongly minimal in $\mathcal{M}$, then by (3) X is very ample, so by Theorem 4.14, X is $K/\mathcal{M}$-full. $\square$

Remark 4.17. The reader may be tempted to replace “almost strongly minimal” with “uncountably categorical” in the statement of Theorem 4.16. Let us point out, then, that almost strong minimality and unidimensionality are not equivalent for ACF-relics. Indeed, for any algebraically closed field, K, consider the two sorted relic $\mathcal{M}$ defined as follows: we define the two sorts to be $M := K^2$ and $U := K$. We then equip $\mathcal{M}$ with the binary relation V on M interpreted as

$$\{((x, y), (x, z)) : x, y, z \in K\},$$

the map

$$f((x, y), (x, z)) = y - z$$

from V to U, and the full field structure on U. (Essentially, we have defined a K-indexed family of copies of K, each of which has “forgotten where 0 is” but remembers an isomorphism with K after fixing any element to be 0.) Clearly, $\mathcal{M}$ is a K-relic. We leave it as an exercise to the interested reader to verify that $\mathcal{M}$ is uncountably categorical, but not almost strongly minimal.

5. Technical results on strongly minimal structures

Here we prove some useful facts about general strongly minimal structures, without assuming very ampleness or Zilber's trichotomy. Though these are technical results needed for the sequel, we hope they will be seen as interesting in their own right. Throughout this section $\mathcal{M}$ is assumed to be a strongly minimal structure.

We will use the following well-known notions.

Definition 5.1. Let p and q be stationary types over A and B, respectively.

(1) The *Morley product* of p and q, denoted $p \otimes q$, is $\mathrm{tp}(ab/AB)$, where $a \models p$, $b \models q$, $a \perp_A B$, and $b \perp_B Aa$.

(2) Given a positive integer k, the k-th *Morley power* of p, denoted p^k, is the k-fold Morley product of p with itself.

(3) The stationary types p and q are *interalgebraic* if $A = B$ and for $a \models p$, there is $b \models q$ with a and b interalgebraic over A.

5A. ***Sweeping extensions.*** Suppose that we are given a 3-dimensional stationary set X, and a high-dimensional family $\mathcal{C}$ of curves in X covering X generically. A subfamily $\mathcal{C}'$ of $\mathcal{C}$ (over additional parameters) could have high dimension, but concentrate on a forking subset of X. In the present section, we develop a notion of largeness of subfamilies aimed at avoiding precisely such situations.

The following notion appeared in an equivalent form in the first author's Ph.D. thesis, and later appeared implicitly in [Castle 2023]. Below, if p and q are complete types, by $p \subset q$ we mean that q is an extension of p.

Definition 5.2. Let $p \subset q$ be stationary types. We say that q is a *k-sweeping* extension of p if $p^k \subset q^k$.

Definition 5.3. Let $A \subset B$ and let p be a stationary type over B. We say that p *k-sweeps* over A if p is a k-sweeping extension of its restriction to $\mathrm{acl}(A)$.

Definition 5.4. If $A \subset B$ and a is a tuple, we say that a *k-sweeps* from B to A if $\mathrm{stp}(a/B)$ k-sweeps over A.

Note that if p is a stationary type over A, and q is a k-sweeping extension of p, then for any Morley sequence $\bar{a} = a_1, \ldots, a_k$ in p, there is an A-automorphism conjugate q' of q such that $\bar{a}$ is a Morley sequence in q'. We, thus, think of the conjugates of q as "filling out" p "to order k". In particular, k-sweeping extensions provide a workable measure of "largeness" for families of extensions of a type. Now, given a stationary set X and a family $\mathcal{C}$ of (generically stationary) subsets of X, we can ask whether the generic type of a generic member of the family is a k-sweeping extension of the generic type of X. This gives a stronger, and geometrically more accurate, notion of largeness for families of definable subsets of X, than just the dimension of the family.

The following facts are easily verified:

Fact 5.5. Let p, q, r, s be stationary types, A and B be parameter sets, a and b be tuples, and k be a positive integer.

(1) If $q \supset p$ is nonforking, then $q \supset p$ is k-sweeping.

(2) If $r \supset q$ and $q \supset p$ are k-sweeping, then $r \supset p$ is k-sweeping.

(3) If $q \supset p$ is k-sweeping, q and s are interalgebraic, p and r are interalgebraic, and $s \supset r$, then $s \supset r$ is k-sweeping.

(4) If ab k-sweeps from B to A, then a k-sweeps from B to A.

(5) If a k-sweeps from B to A, and $b \perp_{Aa} B$, then b k-sweeps from B to A.

We also give the following two lemmas, which essentially appeared in Castle's Ph.D. thesis:

Lemma 5.6. *Suppose that a k-sweeps from B to A. Then for any b such that* $\dim(b/A) < k$, *$a \perp_B b$ implies $a \perp_A b$.*

Proof. Assume $a \perp_B b$. Let $\bar{a} = a_1, \ldots, a_k$ be a Morley sequence in $\mathrm{stp}(a/Bb)$. Since $a \perp_B b$, $\bar{a}$ is also a Morley sequence in $\mathrm{stp}(a/B)$, and since a k-sweeps from B to A, $\bar{a}$ is also a Morley sequence in $\mathrm{stp}(a/A)$. Hence $\dim(\bar{a}b/A) \geq k \cdot \dim(a/A)$. Since $\dim(b/A) < k$, this implies $\dim(\bar{a}/Ab) > k \cdot (\dim(a/A) - 1)$. Hence, there is some i with $\dim(a_i/Ab) = \dim(a/A)$. But $\mathrm{tp}(a/Ab) = \mathrm{tp}(a_i/Ab)$, so $\dim(a/Ab) = \dim(a/A)$. $\square$

Lemma 5.7. *Let $p \subset q$ be stationary types over A and B respectively. If* $\dim(q) = \dim(p) - 1$, *then q is a k-sweeping extension of p, where* $k = \dim(\mathrm{Cb}(q)/A)$.

Proof. Let $c = \mathrm{Cb}(q)$. Let $d = \dim(p)$. Let $\bar{a} = (a_1, \ldots, a_k)$ be a Morley sequence in q. So $\dim(c\bar{a}/A) = kd$, counting from left to right. Now if the lemma fails, then $\dim(\bar{a}/A) < kd$, so $\dim(c/A\bar{a}) > 0$. Let c' be an independent realization of $\mathrm{tp}(c/A\bar{a})$ over c. Then by assumption we get (1) $\dim(cc'\bar{a}/A) > kd$, and (2) $\dim(c'/Ac) > 0$. By (2) $c' \neq c$, so each a_i forks with Acc' over Ac, and thus $\dim(\bar{a}/Acc') \leq k(d-2)$. So by (1), $\dim(cc'/A) > 2k$, contradicting that $\dim(c/A) = \dim(c'/A) = k$. $\square$

We conclude this subsection with an existence result for sweeping extensions. It is well known that nonlocal modularity is characterized by the presence of arbitrarily large families of one-dimensional subsets of M^2; see also Conditions (A), (B) and (C) on pp. 85–87 of [Hrushovski 1989]. By Lemma 5.7, this is equivalent to the generic type of M^2 having k-sweeping one-dimensional extensions for arbitrarily large k. Proposition 5.8 says that there is nothing special about one-dimensional subsets of M^2: namely, adopting sweeping as a notion of largeness, every definable set X admits arbitrarily large definable families of m-dimensional subsets, for any $0 < m \leq \dim X$.

Proposition 5.8. *Assume $\mathcal{M}$ is not locally modular. Let $0 < m \le n$ and k be positive integers, and let p be a stationary type over A with* $\dim(p) = n$. *Then there is a k-sweeping extension $q \supset p$ with* $\dim(q) = m$.

Proof. We start with some reductions. If $m = n$, then p is a k-sweeping extension of itself. Now assume $m < n$. By Fact 5.5 (2), it suffices to assume $m = n-1$. Now by weak elimination of imaginaries, there is a nonforking extension $p' \supset p$, say over $A' \supset A$, such that p is interalgebraic with the generic type p'' of M^n over A'.

By nonlocal modularity, there is a strongly minimal plane curve $C \subset M^2$, with canonical base c, such that $\dim(c/A') \ge k$. Let r be the generic type of C over $A'c$, and let $q' = r \otimes s$, where s is the generic type of M^{n-2} over $A'c$. So $\dim(q') = n-1$, and clearly $\mathrm{Cb}(q')$ is also c. Then by Lemma 5.7, and since $\dim(c/A') \ge k$, the extension $q' \supset p''$ is k-sweeping. Let $a \models q'$. So $a \models p''$. By interalgebraicity, there is $b \models p'$ with a and b interalgebraic over A'. Now let $q = \mathrm{stp}(a/A'c)$. By interalgebraicity, $\dim(q) = n-1$, and by Fact 5.5 (3), $q \supset p'$ is k-sweeping. But by Fact 5.5 (1), $p' \supset p$ is also k-sweeping, thus by Fact 5.5 (2), so is $q \supset p$. $\square$

Note that nonlocal modularity is crucial in the above proposition. Indeed, in any stable one-based theory, if $q \supset p$ are stationary types with $q^2 \supset p^2$, then q is a nonforking extension of p. To see this, suppose p is over A, let $c = \mathrm{Cb}(q)$, and let $(a, b) \models q^2$. By one-basedness, $c \in \mathrm{acl}(a) \cap \mathrm{acl}(b)$. But since $a \mathop{\smile\!\!\!\!\!\!\;\mid}_A b$, this gives $c \in \mathrm{acl}(A)$, implying that q does not fork over A.

5B. ***Sets transverse to any family.*** The result of this subsection was inspired by the following question: given a definable family $\{X_t : t \in T\}$ of codimension r subsets of a definable set Z, can one always find an r-dimensional definable set $Y \subset Z$ having finite intersection with each X_t? This type of question arises, for example, when trying to extract a 2-dimensional very ample family of plane curves from a higher dimensional such family (see Proposition 6.1).

In order to answer the question above, it is convenient to study a more general question. Namely, for type-definable sets $X, Y \subset Z$, say that X and Y are *transverse in* Z if $\dim(X \cap Y) \le \dim X + \dim Y - \dim Z$. We wish to investigate whether, given a type-definable family $\{X_t : t \in T\}$ of subsets of a definable set Z, one can find a definable $Y \subset Z$ of any prescribed dimension which is transverse to each X_t in Z.

As stated, it is not always possible to find such Y: indeed, if the prescribed dimension of Y is small enough (but still at least 0), transversality could require Y to be disjoint from each X_t, which is clearly impossible if the X_t cover Z. So at the least, we must allow Y to be nonempty. Conversely, Proposition 5.9 and Corollary 5.10 show that this is, essentially, the only obstacle.

Proposition 5.9. *Assume $\mathcal{M}$ is not locally modular. Fix nonnegative integers i, j, m, k with $i, j \le m$. Let p be a stationary type over A of dimension m. Let*

$\{X_t : t \in T\}$ be a type-definable family of sets over A, with each $\dim(X_t) \le i$. Then there are a set $B \supset A$ and a stationary type $q \supset p$ over B satisfying:

(1) $\dim(q) = j$.

(2) *If $j \ge 1$, then $q \supset p$ is a k-sweeping extension.*

(3) *If $a \models q$ and $a \in X_t$ for some t, then $\dim(a/Bt) \le \max\{0, i + j - m\}$.*

Proof. For ease of notation, we assume throughout that $A = \varnothing$. Note that it is harmless to increase the value of k, so we may assume $k > \dim T$.

We proceed by induction on j. If $j = 0$ then conditions (2) and (3) are automatic, and one can simply take $q = \mathrm{tp}(a/a)$ where $a \models p$.

Now assume $j \ge 1$, and let $r \supset p$ satisfy the proposition with j replaced by $j-1$. Suppose r is over C. Let $c = \mathrm{Cb}(r)$. Since $j - 1 < j \le m$, r is a forking extension of p, thus $\dim(c) \ge 1$. Then, by Proposition 5.8, there is a set B such that $\mathrm{tp}(c/B)$ is minimal and k-sweeps over $\varnothing$. Without loss of generality, B is algebraically closed. Finally, let r' be the nonforking extension of r over BC, and let q be the restriction of r to B. Since B is algebraically closed, q is stationary. We show in the ensuing four claims that q satisfies the requirements of the proposition.

Claim 5.9.1. $\dim(q) = j$.

Proof. Let $a \models r'$. We want to show that $\dim(a/B) = j$. Now since r' is a nonforking extension of r, we have

$$\dim(a/Bc) = \dim(r') = \dim(r) = j - 1.$$

Since $\dim(c/B) = 1$, this implies $j-1 \le \dim(a/B) \le j$. So assume $\dim(a/B) = j-1$. Then r' does not fork over B, and since $c = \mathrm{Cb}(r) = \mathrm{Cb}(r')$, this implies $c \in \mathrm{acl}(B)$, contradicting that $\dim(c/B) = 1$. □

Claim 5.9.2. *The stationary type q is a k-sweeping extension of p.*

Proof. Let $a \models r'$. It is enough to show that a k-sweeps from B to $\varnothing$. Now by assumption, c k-sweeps from B to $\varnothing$. Moreover, since $r' \supset r$ is nonforking, a and B are independent over c. So, by Fact 5.5 (5), a k-sweeps from B to $\varnothing$. □

Item (3) in the proposition will be the most difficult to establish. The key observation is:

Claim 5.9.3. *Let $a \models r'$ and $a \in X_t$ for some t. Then either $c \in \mathrm{acl}(Bt)$ or $i + (j-1) - m \ge 0$.*

Proof. By various instances of additivity, we have

$$\begin{aligned}\dim(c/t) &\le \dim(ac/t) = \dim(a/t) + \dim(c/at)\\ &\le \dim(a/t) + \dim(c/a) = \dim(a/t) + \dim(c) + \dim(a/c) - \dim(a)\\ &\le i + \dim(c) + (j-1) - m.\end{aligned}$$

Rearranging,

$$\dim(c/t) - \dim(c) \leq i + (j-1) - m.$$

Now $\dim(c/B) = 1$ by assumption, so if $c \notin \operatorname{acl}(Bt)$, then $c \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.85em\lower0.4ex\hbox{$\mid$}}_B t$. But by construction, c k-sweeps from B to $\varnothing$, and $k > \dim(T) \geq \dim(t)$. So by Lemma 5.6, we get $c \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.85em\lower0.4ex\hbox{$\mid$}}_\varnothing t$, and the above simplifies to $0 \leq i + (j-1) - m$. $\square$

Finally, we complete the proof of Proposition 5.9 by showing (3):

Claim 5.9.4. *Let $a \models q$ and $a \in X_t$ for some t. Then* $\dim(a/Bt) \leq \max\{0, i+j-m\}$.

Proof. After applying an automorphism fixing B, we may assume $a \models r'$. Let $a't' \models \operatorname{tp}(at/c)$ with $a't' \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.85em\lower0.4ex\hbox{$\mid$}}_c C$. So $a' \models r$ and $a' \in X_{t'}$. Combined with the inductive hypothesis, we get

$$\dim(a/ct) = \dim(a'/ct') = \dim(a'/Ct') \leq \max\{0, i + (j-1) - m\}.$$

We now proceed in two cases, according to the previous claim.

Case 1: Assume $c \in \operatorname{acl}(Bt)$. Then

$$\dim(a/Bt) \leq \dim(a/ct) \leq \max\{0, i + (j-1) - m\} \leq \max\{0, i + j - m\}.$$

Case 2: Assume $i+(j-1)-m \geq 0$. Then $\max\{0, i+(j-1)-m\} = i+(j-1)-m$, so $\dim(a/ct) \leq i + (j-1) - m$. But since $\dim(c/B) = 1$, we then have

$$\dim(a/Bt) \leq \dim(a/ct) + 1 = i + j - m,$$

as claimed. $\square$

This finishes the proof of the proposition. $\square$

Combining Proposition 5.9 with compactness, we obtain:

Corollary 5.10. *Let i, j, m be integers with $0 \leq i, j \leq m$. Let Z be a definable set of dimension m, and let $\{X_t : t \in T\}$ be a definable family of subsets of Z, with each* $\dim(X_t) \leq i$. *Then there is a definable set $Y \subset Z$ of dimension j such that for each $t \in T$, we have* $\dim(X_t \cap Y) \leq \max\{0, i + j - m\}$.

6. Very ample families of a prescribed dimension

As we have seen, very ampleness in strongly minimal structures is equivalent to the existence of very ample families of plane curves (Corollary 3.21). In applications, it may be convenient to have access to such families of a prescribed dimension (as is the case with nonlocal modularity, where for any $k > 1$ the existence of any k-dimensional family of plane curves implies the existence of families of all dimensions). In the present section, we show that this can always be achieved. Namely, we first use Proposition 5.9 to show that, given a k-dimensional very ample family of plane curves, we can extract an l-dimensional very ample subfamily for any $2 \leq l \leq k$. We then adapt the usual proof from the nonlocally modular setting (using composition and a field configuration) to construct very ample families of arbitrarily large dimension. While the general proof structure is the same as in the nonlocally modular case, the need to preserve very ampleness will make each stage of the argument more delicate. In particular, once a field (say K) is interpreted in our set (say X), some care is needed to examine the precise relationship between X and K.

Once we have achieved the result described above, we end the section with a short application, in which (as mentioned in the introduction to the paper) we characterize very ampleness in terms of definable pseudoplanes.

6A. *Going down.* Here we prove that very ample families of plane curves admit very ample subfamilies of any prescribed dimension. The proof amounts to an application of Proposition 5.9.

Proposition 6.1. *Let X be strongly minimal and A-definable. Let C be a strongly minimal plane curve in X, let $c = \mathrm{Cb}(C)$, and let $m = \dim(c/A)$. Assume that C is very ample over A. Then for all $2 \leq j < m$, there is a set $B \supset A$ such that $\dim(c/B) = j$ and C is very ample over B.*

Proof. Without loss of generality, assume A is algebraically closed, as this does not affect any of the relevant computations. Let $p = \mathrm{tp}(c/A)$. So p is stationary of dimension m. Let a be generic in C over Ac, and let $T = X^2 \times X^2 \setminus \Delta$, where Δ is the diagonal $(v, v) \in X^2 \times X^2$. Let $\{X_t : t \in T\}$ be the A-type-definable family (in the variables v, v', z) defined as follows: for $t = (v, v')$, set $z \in X_t$ if $\mathrm{tp}(vz/A) = \mathrm{tp}(v'z/A) = \mathrm{tp}(ac/A)$. Then X_t is the set of A-automorphic images of C, passing generically through v and v'. It follows from the very ampleness of C over A that $\dim(X_t) \leq m - 2$ for all $t \in T$ (see Lemma 3.11).

Now let $i = m-2$ and $k = 1$, and let $B \supset A$ and q be as provided by Proposition 5.9 (for j as given in the statement of the current proposition). After applying an automorphism, we may assume that $c \models q$. So $\dim(c/B) = j$. We will show that C is very ample over B, proving the proposition.

To show very ampleness, let v, v' be distinct generics of C over c. By Lemma 3.11, it is enough to show that $\dim(c/Bvv') \leq j - 2$. But by definition we have $c \models q$ and $c \in X_{(v,v')}$, so $\dim(c/Bvv') \leq \max\{0, i + j - m\}$. But $i = m - 2$, so $i + j - m = j - 2$. Thus $\dim(c/Bvv') \leq \max\{0, j - 2\}$, and since $j \geq 2$, this implies $\dim(c/Bvv') \leq j - 2$. □

6B. *Going up.* Our next goal is to show that if X is a very ample strongly minimal set, then X has arbitrarily large very ample families of plane curves. Before proceeding, we need the following well-known lemma[5]:

[5] The argument goes back to the proof that nonlocal modularity implies the existence of faithful families of plane curves or arbitrarily large dimension. Similar arguments can be found in [Hasson and Sustretov 2017, Lemma 4.18] and [Eleftheriou et al. 2021, §3]

Lemma 6.2. *Let X be an A-definable strongly minimal set, and let C, D, and $E \subset D \circ C$ be strongly minimal plane curves in X, with canonical bases c, d, and e, respectively. Assume that $c \downarrow_A d$ and $\dim(e/A) \leq \dim(c/A) = \dim(d/A) = k$ for some $k \geq 2$. Then $k \in \{2, 3\}$, and there is a definable strongly minimal expansion of an algebraically closed field which is internal to X.*

Proof. Take $(x, z) \in E$ generic and y such that $(x, y) \in C$ and $(y, z) \in D$. The assumption implies that $\{c, d, e, x, y, z\}$ is a group configuration; see [Eleftheriou et al. 2021, Lemma 3.18, Lemma 3.20] for the details. By a result of Hrushovski (see [Poizat 2001, Theorem 3.27]) this implies that $k \leq 3$, and if $\dim(k) \geq 2$ (which holds by assumption), a strongly minimal field is definable in X^{eq}. □

We now give the main result of this subsection.

Proposition 6.3. *Suppose X is strongly minimal, A-definable, and very ample. Then for any $k \geq 2$ one can find a set $B \supset A$, and a strongly minimal plane curve C in X, such that $\dim(\mathrm{Cb}(C)/B) \geq k$ and C is very ample in X over B.*

Proof. For simplicity of notation, we assume that $A = \varnothing$. Assume the proposition fails. By very ampleness, there is a strongly minimal plane curve C in X (with canonical base c, say) which is very ample in X over some set B. Fix such (C, c, B) so that the value $k := \dim(c/B)$ is maximal (that this is possible follows by the assumed failure of the proposition). We may assume moving forward that $B = \varnothing$. Note also that $k \geq 2$ by Lemma 3.8.

Let $d \models \mathrm{tp}(c/B)$, with $c \downarrow d$. So $\dim(d) = k$, and $d = \mathrm{Cb}(D)$ for some strongly minimal plane curve D in X which is also very ample in X over $\varnothing$. Let $E \subset D \circ C$ be strongly minimal. Then, by Lemma 3.23 and Corollary 3.14, E is also very ample over $\varnothing$. So by the choice of k, letting $e = \mathrm{Cb}(E)$, we have $\dim(e) \leq k$. Thus, by Lemma 6.2, $k \in \{2, 3\}$ and there is a strongly minimal expansion $(K, +, \cdot, \ldots)$ of an algebraically closed field internal to X.

Since K is internal to X, clearly X and K are nonorthogonal. So by Corollary 4.2, there is a definable embedding of X into some K^n. Let us choose, among all definable embeddings of cofinite subsets of X into any K^n, one which minimizes the value n. So, replacing X with a cofinite subset if necessary, we assume that $X \subset K^n$ and no cofinite subset of X embeds into any smaller power of K. Without loss of generality, we will assume that K and X are $\varnothing$-definable.

In what follows, it will be convenient to introduce the following notation: if $S \subseteq K^m$ is definable, we let S^{aff} be the intersection of all affine linear subspaces of K^n which contain a large subset of S. It is easy to see that S^{aff} is itself affine linear and contains a large subset of S, and moreover that S^{aff} is definable over any set of parameters defining S. Note also that by the choice of n, we have $X^{\mathrm{aff}} = K^n$ (since any proper affine linear subspace of K^n is in definable bijection with a lower power of K).

Let $\mathcal{H} = \{H_t : t \in T\}$ be the family of hyperplanes in K^{2n}. So $\dim(T) = 2n$. We note the following basic observation, for future reference:

Claim 6.3.1. *If $D \subseteq K^{2n}$ is a d-dimensional affine linear subspace, then we have* $\dim(\{t \in T : D \subseteq H_t\}) = 2n - d - 1$.

Now let $t \in T$ be generic. Then, by Proposition 3.28, $H_t \cap X^2$ is a very ample (over $\varnothing$) plane curve in X. Let $S \subset H_t \cap X^2$ be strongly minimal. Then, by Corollary 3.14, S is also very ample in X over $\varnothing$. So by the choice of k, letting $s = \mathrm{Cb}(S)$, we have $\dim(s) \leq k \leq 3$, and thus $\dim(s) \leq 3$.

Since S is a strongly minimal component of $H_t \cap X^2$, clearly $s \in \mathrm{acl}(t)$. So $\dim(ts) = \dim(t) = 2n$, and thus $\dim(t/s) \geq 2n - 3$. On the other hand, by definition H_t contains S, so since H_t is affine linear it must also contain S^{aff}. Let $d := \dim(S^{\mathrm{aff}})$; then by Claim 6.3.1, we have $\dim(t/s) \leq 2n - d - 1$. So, combining with $\dim(t/s) \geq 2n - 3$, we get $2n - 3 \leq 2n - d - 1$, and thus $d \leq 2$. Note also that if equality holds (i.e., $d = 2$), we have $2n - 3 \leq \dim(t/s) \leq 2n - 3$, so that $\dim(t/s) = 2n - 3$, and thus $\dim(s) = 3$.

Now, the main point of the proof is the following:

Lemma 6.4. *There is a definable strongly minimal algebraically closed field* $(F, +_F, \cdot_F)$, *whose underlying set F almost coincides with X.*

Proof. First note that if $\dim(X^{\mathrm{aff}}) = 1$, then X almost coincides with X^{aff}, and X^{aff} is in definable bijection with K; in particular, this is enough to prove the lemma (setting $F = X^{\mathrm{aff}}$ with the field structure inherited from K). So we may assume that $n = \dim(X^{\mathrm{aff}}) \geq 2$, and thus (as above) $\dim(s) = 3$. Our goal is to construct a definable transitive group action of a 3-dimensional group on a set almost equal to X, and apply Hrushovski's classification [Poizat 2001, Theorem 3.27] of such actions. We first proceed with the following two claims, which give us the group we will use:

Claim 6.4.1. *Let $\pi : K^{2n} \to K^n$ be either the leftmost or rightmost projection. Then the restriction of π to S^{aff} is surjective.*

Proof. By Remark 3.9, π is finite-to-one on S, and thus $\pi(S)$ contains a cofinite subset of X. By definition, S^{aff} contains a cofinite subset of S, so clearly $\pi(S^{\mathrm{aff}})$ contains a cofinite subset of $\pi(S)$. In particular, $\pi(S^{\mathrm{aff}})$ also contains a cofinite subset of X, and thus contains $X^{\mathrm{aff}} = K^n$ (because projections of affine sets are affine). □

We next use Claim 6.4.1 to show the following:

Claim 6.4.2. *The intersection S^{aff} is the graph of an invertible affine linear map* $L : K^2 \to K^2$.

Proof. By Claim 6.4.1 we have $d \geq n \geq 2$; so since we already had $d \leq 2$, we get $d = n = 2$. Then by Claim 6.4.1 again, each of the two projections $S^{\mathrm{aff}} \to K^2$ is a surjective linear map between two-dimensional affine linear spaces; thus each of these projections is a linear bijection, which implies the claim. □

Let $L : K^2 \to K^2$ be as in Claim 6.4.2. Since S projects almost onto X in both directions, and by the strong minimality of X, it follows that $L(X)$ almost coincides with X. By the strong minimality of S, it moreover follows that S almost coincides with the graph of the restriction of L to X. In particular, it follows easily that s is interdefinable with the canonical parameter of L.

Now let G be the generic stabilizer of X in $\mathrm{AGL}_2(K)$ — that is, the set of all invertible affine linear maps $g : K^2 \to K^2$ such that $g(X)$ almost coincides with X. Clearly, G is a $\varnothing$-definable subgroup of $\mathrm{AGL}_2(K)$. By the above remarks (since s is definable from L), we have $\dim(G) \geq \dim(s) = 3$. Let G^0 be the connected component of G; so $\dim(G^0) \geq 3$ and G^0 is $\varnothing$-definable.

We now have the group for our action; our next goal is to find an appropriate set which is acted on transitively. Now G^0 comes equipped with both an action on K^2, and a generic action on X: so if $g \in G^0$ and $x \in X$ then $g(x) \in K^2$ is always defined, while if g and x are independent generics then $g(x)$ is also a generic of X. Fix $x_0 \in X$ generic and independent of s, and let $G^0(x_0)$ be the orbit of x_0 under all of G^0. Note, then, that $G^0(x_0)$ is definable over x_0. Our aim is to show that $G^0(x_0)$ almost coincides with X, and subsequently build a field structure on $G^0(x_0)$. We will do this via a sequence of claims. To start, we show:

Claim 6.4.3. *The orbit $G^0(x_0)$ almost contains X.*

Proof. Since x_0 is generic over s, there is $y_0 \in X$ such that (x_0, y_0) is generic in S. So $y_0 = L(x_0)$. Now, by genericity, we have $\dim(x_0 y_0/s) = 1$; Note then that $\dim(x_0 y_0/\varnothing)$ cannot also be 1, because then we would have $s \in \mathrm{acl}(\varnothing)$, contradicting $\dim(s) = 3$. So $\dim(x_0 y_0) = 2$. In particular, y_0 is generic in X over x_0. Since $y_0 = L(x_0)$, we see that $y_0 \in G^0(x_0)$, and the claim follows. □

Using Claim 6.4.3, we now show:

Claim 6.4.4. *The orbit $G^0(x_0)$ is one-dimensional, and X is (up to a finite set) one of its strongly minimal components.*

Proof. Let D be the set of $g \in G^0$ with $g(x_0) \in X$. Then D is definable over x_0 and contains all generics of G^0 over x_0, which implies that D is generic in G^0. So there are $g_1, \ldots, g_m \in G^0$ such that the translates $g_i D$ cover G^0. Then $G^0(x_0)$ is contained in the union of the $g_i D(x_0)$, which by definition is contained in the union of the $g_i(X)$. Now each $g_i(X)$ is strongly minimal, which implies that $G^0(x_0)$ is contained in a one-dimensional definable set. The claim then follows by Claim 6.4.3. □

Finally, we show:

Claim 6.4.5. *The orbit $G^0(x_0)$ is strongly minimal, and thus almost coincides with X.*

Proof. As a transitive G^0-set, $G^0(x_0)$ is isomorphic to a coset space G/H, where H is the stabilizer of x_0. It is easy to see that this identification is definable (that is, the usual identification from basic algebra is clearly definable). In this case, the strong minimality of G^0/H follows easily from the fact that G^0 is connected: indeed, given a partition of G^0/H into disjoint definable infinite sets Z_1 and Z_2, the preimages of the Z_i in G^0 would partition G^0 into two generic subsets, contradicting connectedness. □

Now by Claim 6.4.5, there is a definable (transitive by definition) action of G^0 on the strongly minimal set $G^0(x_0)$, and $\dim(G^0) \geq 3$. So by the classification of such actions [Poizat 2001, Theorem 3.27], $\dim(G^0) = 3$ and $G^0(x_0)$ is in definable bijection with $\mathbb{P}^1(F)$ for some definable, strongly minimal, algebraically closed field F. Deleting a point, we then recover $\mathbb{A}^1(F) = F$; then by Claim 6.4.5 again, X is (up to a definable bijection) almost equal to F, proving Lemma 6.4. □

Now by Lemma 6.4, we may assume after changing finitely many points that there is a definable field structure on the set X. It is then an easy exercise to find arbitrarily large very ample families of plane curves (for example, the family of degree d polynomial maps $X \to X$ is very ample for each d). In particular, there are very ample (over $\varnothing$) strongly minimal plane curves $C \subset X^2$ with $\dim(\mathrm{Cb}(C)) > k$, contradicting our initial assumption and thus proving Proposition 6.3. □

The main part of the proof is showing that the conclusion of Proposition 6.1 is true for strongly minimal sets definable in strongly minimal expansions of algebraically closed fields. It may be interesting to know whether a simpler proof of this fact exists in the case of pure algebraically closed fields.

Finally, we now deduce:

Corollary 6.5. *Let X be a very ample strongly minimal set. Then for every $k \geq 2$, there is a faithful, very ample family $\{C_t : t \in T\}$ of plane curves in X where* $\dim(T) = k$.

Proof. Assume X is A-definable. Applying Proposition 6.3 and then Proposition 6.1, we can find a set $B \supset A$ and a strongly minimal plane curve $C \subset X^2$ which is very ample in X over A and satisfies $\dim(\mathrm{Cb}(C)/B) = k$. After editing finitely many points of C and applying compactness, there is then a B-definable faithful family of plane curves in X whose generic members are the conjugates of C over B. The result then follows by applying Proposition 3.19 to this family. □

6C. ***Very ampleness and pseudoplanes.*** We now show, as described in the introduction, that a strongly minimal set X is very ample (i.e., admits a very ample plane curve) if and only if there is a definable pseudoplane on (a large subset of) X^2. Let us begin by formally recalling the definition.

Definition 6.6. A *pseudoplane* consists of sets P and L, and an incidence relation $I \subset P \times L$, satisfying the following:

(1) For each $l \in L$, there are infinitely many $p \in P$ with $(p, l) \in I$.

(2) For each $p \in P$, there are infinitely many $l \in L$ with $(p, l) \in I$.

(3) For each $l \neq l' \in L$, there are only finitely many $p \in P$ with $(p, l), (p, l') \in I$.

(4) For each $p \neq p' \in P$, there are only finitely many $l \in L$ with $(p, l), (p', l) \in I$.

One thinks of P as the set of "points" of a plane, and L as the set of "lines". As mentioned above, already Zilber in his thesis proved the "weak trichotomy theorem", asserting that a strongly minimal structure is not locally modular if and only if it defines a pseudoplane. However, if one wishes to make the natural identification $P := M^2$ (or a large subset thereof), one can only guarantee conditions (1)–(3) in the definition. We now point out that the missing data in order to identify P (generically) with M^2 is precisely very ampleness.

Proposition 6.7. *Let X be strongly minimal. Then the following are equivalent*:

(1) *X is very ample.*

(2) *There is a definable pseudoplane (P, L, I), where P is a generic subset of X^2.*

Proof. First, assume X is very ample. Then by Proposition 6.1, there is a definable faithful very ample family of plane curves $\mathcal{C} = \{C_t : t \in T\}$ in X, where $\dim(T) = 2$. Let $C \subset X^2 \times T$ be the graph of $\mathcal{C}$. Let P be the set of $x \in X^2$ such that the set $C^x \subset T$ has dimension 1. Let L be the set of $t \in T$ such that $C_t \cap P$ is infinite. Let $I = C \cap (P \times L)$. It is easy to check that P is generic in M^2 and L is cofinite in T. It then follows easily that axioms (1) and (2) of Definition 6.6 hold for (P, L, I) (these computations are essentially carried out in Section 2 of [Castle 2023]). To conclude, we note that (3) is a restatement of the faithfulness of $\mathcal{C}$, and (4) is a restatement of the very ampleness of $\mathcal{C}$.

Now assume (P, L, I) is a definable pseudoplane, where P is generic in X^2. We will show that $\{I_l : l \in L\}$ is a two-dimensional very ample family of plane curves. By (3) and (4) in the definition, it will suffice to show that $\dim(L) = 2$ and each $\dim(I_l) = 1$. These follows from the ensuing three claims. Throughout, we will assume (P, L, I) (and X) are $\varnothing$-definable.

Claim 6.7.1. *For all $l \in L$, we have* $\dim(I_l) = 1$.

Proof. By assumption $I_l \subset X^2$ and I_l is infinite, so we need only rule out the case that $\dim(I_l) = 2$. But in this case I_l is generic in X^2, so $\dim(X^2 - I_l) = 1$. By the finiteness of Morley degree, it follows easily (using (3)) that there can be only finitely many distinct $I_{l'}$ other than I_l. In other words, we get that L is finite, which clearly contradicts (2). □

Claim 6.7.2.
$$\dim(L) \leq 2.$$

Proof. Let $l \in L$ be generic, and let p, p' be independent generics in I_l over l. By (4), $\dim(l/pp') = 0$, so $\dim(ppl') = \dim(pp') \leq 4$. But by the previous claim $\dim(pp'/l) = 2$, so it follows that $\dim(l) \leq 2$, and thus $\dim(L) \leq 2$. □

Claim 6.7.3. $$\dim(L) \geq 2.$$

Proof. Let $p \in P$ be generic, and let $l \in I^p$ be generic over p. By (2), we have $\dim(l/p) \geq 1$. Now if $\dim(L) < 2$, we are forced to conclude that $\dim(l/p) = \dim(l) = 1$. Thus, l and p are independent. By assumption $\dim(p) = 2$, so also $\dim(p/l) = 2$. But then $\dim(I_l) \geq 2$, contradicting Claim 6.7.1. □

This completes the proof of the proposition. □

Remark 6.8. The above proof would be fairly straightforward, and included much sooner in the paper, if we knew right away that very ampleness gave us a two-dimensional very ample family of plane curves. The reason we had to wait until now to present the result is that we needed Proposition 6.1 to get such a family.

Before moving on, we point out one more fact that is relevant to pseudoplanes. As we have seen, very ampleness implies nonlocal modularity of strongly minimal structures. In Corollary 6.12, we show, conversely, that nonlocal modularity (in fact, non-one-basedness) implies the existence of a very ample stationary type. This is true for stable theories, and the proof goes through *complete type-definable pseudoplanes*. Recall the following (see [Pillay 1996, §4, Definition 1.6]):

Definition 6.9. A *complete-type-definable pseudoplane* consists of a complete type of a pair of (potentially imaginary) tuples $p = \mathrm{tp}(b, c)$ such that

(1) $b \notin \mathrm{acl}(c)$.

(2) $c \notin \mathrm{acl}(b)$.

(3) If $c' \neq c$ and $\mathrm{tp}(bc) = \mathrm{tp}(bc')$, then $b \in \mathrm{acl}(cc')$.

(4) If $b' \neq b$ and $\mathrm{tp}(bc) = \mathrm{tp}(b'c)$, then $c \in \mathrm{acl}(bb')$.

The following is proved in [Pillay 1996, §4, Lemma 1.7]:

Fact 6.10. A stable theory is one-based if and only if it does not admit a complete-type-definable pseudoplane.

Remark 6.11. Fact 6.10 need not be true if we do not insist that a type-definable pseudoplane is concentrated on a complete type. Indeed, the existence of a (type) definable pseudoplane does not imply the existence of a complete type definable pseudoplane, and consequently the existence of a (type) definable pseudoplane in itself does not contradict one-basedness. Let X be an infinite set equipped with a definable surjection, $f : X \to X$ with all fibers infinite. Construct a pseudoplane (P, L, I) by setting $X = P = L$ and $I(p,l)$ if $f(p) = l$ or $f(l) = p$. Clearly, there is no complete type-definable pseudoplane contained in (P, L, I). We leave it as an exercise to the interested reader to find stable one-based structures where such definable pseudoplanes can be constructed.

We now show:

Corollary 6.12. *A stable theory admits a very ample nonalgebraic stationary type if and only if it is not one-based.*

Proof. Let T be a stable theory. First suppose T is one-based. We show that no nonalgebraic stationary type is very ample. To see this, let $p = \mathrm{tp}(a/c)$ be stationary, where $c = \mathrm{Cb}(p)$. Since p is not algebraic, there is $b \models p$ with $b \neq a$. Now by one-basedness, $c \in \mathrm{acl}(a)$, so automatically $c \mathop{\smile\!\!\!\!\!\mid}_a b$, which shows that p is not very ample.

Now suppose T is not one-based. By Fact 6.10, there is a complete-type-definable pseudoplane $\mathrm{tp}(bc)$. Let $d = \mathrm{Cb}(\mathrm{stp}(b/c))$. Clearly, $d \in \mathrm{acl}(c)$. On the other hand, by Definition 6.9, any c-conjugate of $\mathrm{tp}(b/c)$ is not parallel to $\mathrm{tp}(b/c)$. Thus $c \in \mathrm{dcl}(d)$, so in particular, c is interalgebraic with d.

We now claim that $\mathrm{tp}(b/d)$ is very ample. To see this, let b' be any realization which is distinct from b'. We want to show that b' forks with d over b; by interalgebraicity, it is enough to show that b' forks with c over b. By symmetry, it is equivalent to show that $r = \mathrm{tp}(c/bb')$ is a forking extension of $q = \mathrm{tp}(c/b)$. But this is clear, since r is algebraic and q is not. □

7. Very ampleness in strongly minimal groups

In this final section, we give some results related to very ampleness in strongly minimal groups. We first prove that nonlocally modular strongly minimal groups admit nonaffine plane curves (a result that has been long assumed but does not exist in writing). We then use this result to show that Question 3.30 has a positive answer for groups, and that nonlocally modular expansions of divisible groups are already very ample; finally, we apply the very ampleness of divisible groups to characterize fullness of strongly minimal ACF-relics with definable divisible group structures, answering, in particular, an old question on expansions of the multiplicative group.

7A. ***Existence of nonaffine plane curves.*** Hrushovski and Pillay [1987] show that a stable group $\mathcal{G}$ is one-based if and only if every definable subset of G^n (any n) is *affine*, i.e., a boolean combination of cosets of definable subgroups of G^n. Restricted to strongly minimal (expansions of) groups, it has been well known among experts that non-one-basedness (equivalently, nonlocal modularity) even implies the existence of a nonaffine plane curve. While this result is often used in practice, there does not seem to exist a full treatment anywhere in writing[6]. It can, however, be deduced by a particular application of Proposition 5.8. Thus, we dedicate the present section to proving this result as a service to the community.

Throughout this subsection, we fix $\mathcal{G}$, a strongly minimal expansion of a group. We will say that a definable set $X \subseteq G^n$ is *affine* if it is a (finite) boolean combination

[6]The result is claimed in [Kowalski and Randriambololona 2016, Proposition 4.2], but there is a mistake in the proof that we do not see how to bridge.

of cosets of definable subgroups of G^n; and *almost affine* if it is almost equal to an affine set. Note that an affine stationary set is, up to a small correction, a coset of a definable group. This suggests the following definition: a stationary type in some G^n is *affine* if it is the generic type of a stationary affine definable set (equivalently, the generic type of a coset of a connected group). Note that affine sets are almost affine, and the converse holds for one-dimensional sets (since finite sets are affine); more generally, note that arbitrary finite Boolean combinations of affine sets are affine.

Notation 7.1. For any nonempty definable set $X \subset G^n$, we let $\mathrm{Stab}(X)$ denote the set of $g \in G^n$ such that $g + X$ almost coincides with X; so $\mathrm{Stab}(X)$ is a subgroup of G^n which is definable over any set of parameters defining X.

We will need the following, which is well known and is left as an easy exercise to the reader (see [Eleftheriou et al. 2021, Lemma 3.10] and Example 3.18 and, for (2), see also the proof of Lemma 7.4 and Lemma 7.6 below):

Lemma 7.2. *Let $\mathcal{G} = (G, +, \dots)$ be a strongly minimal expansion of a group, and let $C \subset G^2$ be a strongly minimal plane curve.*

(1) *The following are equivalent*:
   (a) *The family of translates $\{C + t : t \in T^2\}$ is faithful.*
   (b) *The family of translates $\{C + t : t \in T^2\}$ is very ample.*
   (c) *The group* $\mathrm{Stab}\,(C)$ *is trivial.*

(2) *The following are also equivalent*:
   (a) *C is affine.*
   (b) $\mathrm{Stab}(C)$ *is infinite.*

We will show:

Theorem 7.3. *Let $\mathcal{G} = (G, +, \dots)$ be a strongly minimal expansion of a group. Then the following are equivalent*:

(1) *$\mathcal{G}$ is locally modular.*

(2) *Every definable plane curve in $\mathcal{G}$ is affine.*

(3) *Every definable one-dimensional set $X \subset G^n$, for any n, is affine.*

(4) *Every stationary type in G^n, for every n, is affine.*

(5) *Every definable subset of each G^n is almost affine.*

(6) *Every definable subset of each G^n is affine.*

We will be brief, as many parts of the above proof are standard, and follow from either well known arguments or the main result in [Hrushovski and Pillay 1987]. Our contribution will be the quick proof of the implication (3) $\Longrightarrow$ (4) using sweeping extensions of stationary types. In fact there is also a rather easy direct proof of the result of Hrushovski and Pillay [1987] restricted to the strongly minimal case (using only the equivalence of local modularity and one-basedness). So, for completeness, we now give this proof (so that our proof of Theorem 7.3 does not depend on [Hrushovski and Pillay 1987]).

Lemma 7.4. *The following are equivalent*:

(1) *$\mathcal{G}$ is locally modular.*

(2) *If $\{X_t : t \in T\}$ is a nonempty faithful family of definable subsets of a definable set Y, each of dimension k, then* $\dim(T) \leq \dim(Y) - k$.

(3) *Every definable set $X \subset G^n$ (for all n) is affine.*

Moreover, if (1)–(3) *hold, then every definable subgroup $H \subset G^n$ (for all n) is* $\mathrm{acl}(\varnothing)$*-definable.*

Proof. (1) $\Longrightarrow$ (2): Assume $\mathcal{G}$ is locally modular, and let $\{X_t : t \in T\}$, Y, and k be as in (2). Without loss of generality $\{X_t\}$ and Y are $\varnothing$-definable. Let $t \in T$ be generic and $x \in X_t$ generic over t. Then $\dim(xt) = \dim(T) + k$, and by one-basedness (equivalently local modularity) $t \in \mathrm{acl}(x)$, so $\dim(x) = \dim(T) + k$, thus $\dim(Y) \geq \dim(T) + k$, which is equivalent to (2).

(2) $\Longrightarrow$ (3): Let $X \subset G^n$ be stationary of dimension k, without loss of generality $\varnothing$-definable. Applying (2) shows that the family of translates of X is at most $(n-k)$-dimensional (i.e., a generic translate of X has canonical base of dimension at most $n-k$). It follows that $\dim(\mathrm{Stab}(X)) \geq k$, which easily implies that X is almost equal to a coset of $\mathrm{Stab}(X)$. By induction on Morley rank and degree, the conclusion follows for all X.

(3) $\Longrightarrow$ (1): Let S be any strongly minimal plane curve; we show that $\dim(\mathrm{Cb}(S)) \leq 1$, implying local modularity. Now we can write $S = C_{t_0}$ for some generic $t_0 \in T$, where $\mathcal{C} = \{C_t : t \in T\}$ is a $\varnothing$-definable (not necessarily faithful) family of plane curves, and $T \subset G^n$ for some n. Absorbing more constants to the language, if needed, we may assume T is stationary. It follows that the total space $C = \{(x,t) : x \in C_t\}$ of $\mathcal{C}$ is stationary, so (by (3)) almost coincides with a coset H of some connected definable subgroup of G^{n+2}. Note that (since H is a coset) there is some (necessarily $\varnothing$-definable) subgroup $K \leq G^2$ such that each nonempty fiber $H_t \subset G^2$ is a coset of K. In particular, since $t_0 \in T$ is generic, it follows that $S = C_{t_0}$ almost coincides with a coset of K. Thus $\mathrm{Cb}(S)$ is definable over a single element of the one-dimensional group G^2/K, which gives $\dim(\mathrm{Cb}(S)) \leq 1$, as desired.

So we have shown that (1)–(3) are equivalent. Now suppose $\mathcal{G}$ is locally modular. If there is a connected definable subgroup $H \subset G^n$ for some n (say of dimension k) which is not $\mathrm{acl}(\varnothing)$-definable, then there is a $\varnothing$-definable infinite faithful family $\{H_t : t \in T\}$ of k-dimensional subgroups of G^n, whose generic members are connected. Passing to a subfamily and adding parameters if necessary, we may assume that $\dim(T) = 1$. Then the family $\{X_s : s \in S\}$ of all cosets of all H_t is a $\varnothing$-definable faithful $(n-k+1)$-dimensional family of k-dimensional subsets of G^n, which contradicts (2). □

Before proceeding, let us also isolate the following two well-known and easy facts:

Lemma 7.5. *Let $\mathcal{H}$ be a group of Morley rank* 1 *definable in an ω-stable structure. Then any definable subset of $\mathcal{H}$ is affine.*

Proof. Let $S \subseteq \mathcal{H}$ be any definable set, and let $H_1, \dots, H_k$ be the cosets of H^0 (the definable connected component of H). Then each H_i is strongly minimal, so $S \cap H_i$ is either finite or cofinite for all i. Thus S has finite symmetric difference with $\cup\{H_i : \mathrm{RM}(S \cap H_i) = 1\}$. Since the right-hand side is an affine set and the symmetric difference is finite (therefore also affine), the lemma follows. □

Lemma 7.6. *Let p be a stationary type in G^n for some n. Then the following are equivalent*:

(1) *The stationary type p is affine.*

(2) *If $(x, y, z) \models p^3$, then $x + y - z \models p$.*

Proof. It is clear that (1) implies (2), by writing p as the generic type of a coset. Now assume (2) holds, and write p as the generic type of a stationary definable set S. Fix $z \in S$ generic over the parameters defining S. Then (2), equivalently stated, gives that $S - z$ is almost contained in $\mathrm{Stab}(S)$ (since given x, y, z as in (2) it is clear that $(x - z) \in \mathrm{Stab}(S)$). Thus, S is almost contained in $\mathrm{Stab}(S) + z$. On the other hand, since z is generic in S it follows by definition that $\mathrm{Stab}(S) + z$ is almost contained in S, so that in fact $S \sim \mathrm{Stab}(S) + z$. So p is equivalently the generic type of $\mathrm{Stab}(S) + s$, and thus (1) holds. □

Let us now proceed with the proof of Theorem 7.3:

Proof. The implications (1) $\implies$ (2) and (6) $\implies$ (1) are contained in Lemma 7.4. The implication (4) $\implies$ (5) is clear, and (5) $\implies$ (6) is immediate by induction on Morley rank. We show (2) $\implies$ (3) and (3) $\implies$ (4).

(2) $\implies$ (3): We have to prove that if every definable plane curve is affine, then so is every definable one-dimensional set. Let $C \subset G^n$ be definable of dimension 1. Since a finite union of affine sets is affine, we may assume C is strongly minimal. Since C is infinite, there is a projection $\pi : G^n \to G$ with cofinite image in G. Without loss of generality $\pi = \pi_1$ (the leftmost projection). Now for $i = 2, \dots, n$, let $C_i = \pi_{1i}(C) \subset M^2$ (the image of C in the first and i-th coordinates). Then C_i is a strongly minimal plane curve, so by (2) almost coincides with a strongly minimal coset H_i. Deleting finitely many points from C if necessary, we may assume $C_i \subset H_i$ for all i.

Using that H_i is a coset and $\pi_1(C)$ is cofinite in G, it follows that $\pi_1(H_i) = G$. So for each i, there is an element $c_i \in G$ with $(0, c_i) \in H_i$. Then replacing C with $C - (0, c_2, \ldots, c_n)$ if necessary, we may assume each H_i is in fact a subgroup of G^2.

Now let $H = \{(x_1, \ldots, x_n) : (x_1, x_i) \in H_i \text{ for } i = 2, \ldots, n\}$. Then $C \subset H$, and it is easy to verify that H is a one-dimensional definable subgroup of G^n. So by Lemma 7.5, C is affine.

(3)$\Longrightarrow$(4): We have to show that if every definable one-dimensional set is affine, then every stationary type is affine. If $\mathcal{G}$ is locally modular this is immediate by Lemma 7.4. So assume $\mathcal{G}$ is not locally modular. The result now follows readily from the existence of 3-sweeping extensions: Let p be a stationary type in G^n for some n, say over A. If $\dim(p) = 0$, p is clearly affine, so assume $\dim(p) \geq 1$. Then by Proposition 5.8, p has a 3-sweeping stationary extension $q \supset p$ with $\dim(q) = 1$. It follows from (3) that q is affine. Now let $(x, y, z) \models q^3$, so by Lemma 7.6, $x + y - z \models q$. Since $q \supset p$ is 3-sweeping, we get $(x, y, z) \models p^3$ and $x + y - z \models p$. Then by Lemma 7.6 again, p is affine. □

7B. ***Very ampleness in groups.*** We now apply Theorem 7.3 to show that a nonlocally modular strongly minimal expansion of a group G admits a finite subgroup H such that the quotient G/H is very ample. We conclude that nonlocally modular expansions of divisible groups are already very ample.

Theorem 7.7. *Let $\mathcal{G} = (G, +, \ldots)$ be a nonlocally modular strongly minimal expansion of a group. Then*:

(1) *There is a finite subgroup $H \leq G$ such that G/H is very ample. In particular, $\mathcal{G}$ admits a very ample sort.*

(2) *If $\mathcal{G}$ is divisible, then $\mathcal{G}$ itself is very ample.*

Proof. First we prove (1). By Theorem 7.3, there is a strongly minimal plane curve $C \subset G^2$ which is not affine. Then by Lemma 7.2, the group $\operatorname{Stab}(C) \leq G^2$ is finite. Let $H \leq G$ be a finite subgroup such that $\operatorname{Stab}(C) \leq H^2$ (for example, H could be the group generated by all coordinates of elements of $\operatorname{Stab}(C)$). Let C/H be the image of C in $(G/H)^2$, by applying the projection $G \to G/H$ to both coordinates. It follows easily that, in the strongly minimal group (G/H), the set C/H is a strongly minimal plane curve with trivial stabilizer. So by Lemma 7.2 again, G/H is very ample.

Now assume further that G is divisible; we prove (2). Let H be as in (1), and let $h = |H|$. Then the map $xH \mapsto h \cdot x$ is a well-defined map from G/H to G, and is clearly definable. Moreover, by divisibility this map is surjective with finite fibers. Thus, we are in the situation of Lemma 3.24 with $X = G/H$ and $Y = G$. So since G/H is very ample, the lemma implies that so is G. □

7C. *Expansions of algebraic groups.* In the final subsection, we apply Theorem 7.7 to characterize ACF-definable expansions of one-dimensional divisible algebraic groups. Our motivation is as follows. Suppose $(K, +, \times)$ is an algebraically closed field, and let K^{lin} denote the structure $(K, +, \{\lambda_a : a \in K\})$, where λ_a is the map $x \mapsto ax$. Martin [1987] makes the following two conjectures:

(1) There are no intermediate structures between $(K, \times)$ and $(K, +, \times)$.

(2) If $\text{char}(K) = 0$ then there are no intermediate structures between K^{lin} and $(K, +, \times)$.

Conjecture (2) was proved by Marker and Pillay [1990]; however, it seems that (1) has not since been addressed. Our result in this subsection will in fact be a more general statement implying both (1) and (2). Namely, we show:

Theorem 7.8. *Let K be an algebraically closed field, and let $(G, \cdot)$ be the group of K-points of a one-dimensional divisible algebraic group over K. Let G^{lin} denote the structure endowing G with the group operation and all of its endomorphisms* (*as an algebraic group*), *and let G^{Zar} denote the full K-induced structure on G. Then there are no intermediate structures between G^{lin} and G^{Zar}.*

Before proceeding with the proof, let us note:

Remark 7.9. Since every endomorphism of the additive group is a scaling, and every endomorphism of the multiplicative group is a power map (thus definable from the group operation alone), we obtain (1) and (2) above. Moreover, note that Theorem 7.8 also applies to elliptic curves over K.

Proof. Let $\mathcal{G} = (G, \cdot, \dots)$ be a reduct of G^{Zar} which properly expands G^{lin}.

Claim 7.9.1. *There is a nonaffine $\mathcal{G}$-definable set $X \subset G^n$ for some n.*

Proof. It suffices to show that every connected definable subgroup of G^n is definable in G^{lin}. So, let $H \leq G^n$ be a connected definable subgroup, say of dimension d. So H is stationary, which implies there is an almost finite-to-one projection $H \to G^d$. Since H is a subgroup, this implies that $H \to G^d$ is everywhere finite-to-one, and the fibers of $H \to G^d$ are cosets of a finite group, say $K \leq H$. Let $k = |K|$, and let kH be the image of H under scaling by k. Then the projection $kH \to G^d$ is the graph of a definable function $f : G^d \to G^{n-d}$. Since kH is a subgroup, f is a homomorphism. In particular, each coordinate component of f (i.e., $G^d \to G^{n-d} \to G$) is a definable homomorphism from G^d to G, and each of those is built of d G^{lin}-definable G-endomorphisms, so is itself G^{lin}-definable. Thus f is G^{lin}-definable, and therefore so is kH.

Finally, let H' be the preimage of kH under scaling by K; so $H \leq H'$. Note that by divisibility, scaling by k is finite-to-one on G^n; thus $\dim(H) = \dim(kH) = \dim(H')$. Then, since H is connected, it must be the connected component $(H')^0$. In particular,

H is the image of H' under scaling by l for some l. Now to complete the proof, recall that kH is G^{lin}-definable; thus so is H' by definition, and thus (scaling by l) so is H. $\square$

Now by the claim and the main result in [Hrushovski and Pillay 1987] (or Lemma 7.4 if desired), it follows that $\mathcal{G}$ is not locally modular. Then, by divisibility and Theorem 7.7, $\mathcal{G}$ is very ample. Finally, by [Hasson and Sustretov 2017] $\mathcal{G}$ satisfies the Zilber trichotomy; so by Theorem 4.14, $\mathcal{G}$ is full in K. In other words, $\mathcal{G}$ is interdefinable with G^{Zar}, as desired. $\square$

Acknowledgement

We would like to thank E. Hrushovksi for clarifications regarding the definition of very ampleness in [Hrushovski and Zilber 1996].

References

[Abu Saleh and Peterzil 2023] H. Abu Saleh and Y. Peterzil, "Additive reducts of real closed fields and strongly bounded structures", *Model Theory* **2**:2 (2023), 381–404. MR

[Baldwin 1973] J. T. Baldwin, "α_T is finite for $\aleph_1$-categorical T", *Trans. Amer. Math. Soc.* **181** (1973), 37–51. MR

[Castle 2023] B. Castle, "Zilber's restricted trichotomy in characteristic zero", *J. Amer. Math. Soc.* (online publication October 2023).

[Castle et al. 2024] B. Castle, A. Hasson, and J. Ye, "Zilber's trichotomy in Hausdorff geometric structures", preprint, 2024. arXiv 2405.02209

[Eleftheriou et al. 2021] P. E. Eleftheriou, A. Hasson, and Y. Peterzil, "Strongly minimal groups in o-minimal structures", *J. Eur. Math. Soc.* **23**:10 (2021), 3351–3418. MR Zbl

[Hasson and Sustretov 2017] A. Hasson and D. Sustretov, "Incidence systems on Cartesian powers of algebraic curves", preprint, 2017. To appear in *J. Eur. Math. Soc.* arXiv 1702.05554

[Hrushovski 1989] E. Hrushovski, "Unidimensional theories: An introduction to geometric stability theory", pp. 73–103 in *Logic Colloquium '87* (Granada, 1987), edited by H.-D. Ebbinghaus et al., Stud. Logic Found. Math. **129**, North-Holland, Amsterdam, 1989. MR Zbl

[Hrushovski 1993] E. Hrushovski, "A new strongly minimal set", *Ann. Pure Appl. Logic* **62**:2 (1993), 147–166. MR Zbl

[Hrushovski and Pillay 1987] E. Hrushovski and A. Pillay, "Weakly normal groups", pp. 233–244 in *Logic colloquium '85* (Orsay, 1985), edited by C. Berline et al., Stud. Logic Found. Math. **122**, North-Holland, Amsterdam, 1987. MR Zbl

[Hrushovski and Zilber 1996] E. Hrushovski and B. Zilber, "Zariski geometries", *J. Amer. Math. Soc.* **9**:1 (1996), 1–56. MR Zbl

[Kowalski and Randriambololona 2016] P. Kowalski and S. Randriambololona, "Strongly minimal reducts of valued fields", *J. Symb. Log.* **81**:2 (2016), 510–523. MR Zbl

[Loveys 2004] J. Loveys, "Linear reducts of the complex field", *Notre Dame J. Formal Logic* **45**:3 (2004), 161–190. MR Zbl

[Marker 2002] D. Marker, *Model theory*, Graduate Texts in Mathematics **217**, Springer, New York, 2002. MR Zbl

[Marker and Pillay 1990] D. Marker and A. Pillay, "Reducts of $(\mathbf{C}, +, \cdot)$ which contain $+$", *J. Symbolic Logic* **55**:3 (1990), 1243–1251. MR Zbl

[Martin 1988] G. A. Martin, "Definability in reducts of algebraically closed fields", *J. Symbolic Logic* **53**:1 (1988), 188–199. MR Zbl

[Pillay 1996] A. Pillay, *Geometric stability theory*, Oxford Logic Guides **32**, Clarendon, New York, 1996. MR Zbl

[Poizat 1988] B. Poizat, "MM. Borel, Tits, Zil'ber et le général nonsense", *J. Symbolic Logic* **53**:1 (1988), 124–131. MR Zbl

[Poizat 2001] B. Poizat, *Stable groups*, Mathematical Surveys and Monographs **87**, American Mathematical Society, Providence, RI, 2001. MR Zbl

[Rabinovich 1993] E. D. Rabinovich, *Definability of a field in sufficiently rich incidence systems*, QMW Maths Notes **14**, Queen Mary and Westfield College School of Mathematical Sciences, London, 1993. MR Zbl

[Tent and Ziegler 2012] K. Tent and M. Ziegler, *A course in model theory*, Lecture Notes in Logic **40**, Association for Symbolic Logic, La Jolla, CA, 2012. MR Zbl

[Zilber 1993] B. Zilber, *Uncountably categorical theories*, Translations of Mathematical Monographs **117**, American Mathematical Society, Providence, RI, 1993. MR Zbl

[Zilber 2010] B. Zilber, *Zariski geometries: Geometry from the logician's point of view*, London Mathematical Society Lecture Note Series **360**, Cambridge University Press, 2010. MR Zbl

Received 7 Dec 2022. Revised 7 Aug 2023.

BENJAMIN CASTLE:

bcastle@math.berkeley.edu
Department of Mathematics, Ben Gurion University of the Negev, Be'er-Sheva, Israel

ASSAF HASSON:

hassonas@math.bgu.ac.il
Department of Mathematics, Ben Gurion University of the Negev, Be'er-Sheva, Israel

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.259

A model theory for meromorphic vector fields

Rahim Moosa

Dedicated to Boris Zilber on the occasion of his 75th birthday.

Motivated by the study of meromorphic vector fields, a model theory of "compact complex manifolds equipped with a generic derivation" is here proposed. This is made precise by the notion of a *differential* CCM-*structure*. A first-order axiomatisation of existentially closed differential CCM-structures is given. The resulting theory, DCCM, is a common expansion of the theories of differentially closed fields and compact complex manifolds. A study of the basic model theory of DCCM is initiated, including proofs of completeness, quantifier elimination, elimination of imaginaries, and total transcendentality. The finite-dimensional types in DCCM are shown to be precisely the generic types of meromorphic vector fields.

1. Introduction

The model-theoretic approach to systems of (ordinary) algebraic differential equations is via the first-order theory of differentially closed fields in characteristic zero (DCF_0). Such systems of equations, at least in the autonomous case when the equations have constant parameters, can be presented geometrically as algebraic vector fields; namely, a projective algebraic variety X equipped with a rational section $v : X \to TX$ to the tangent space. In fact, the finite-dimensional fragment of DCF_0 essentially coincides with the birational geometry of algebraic vector fields. (See, for example, [Moosa 2022] for an exposition of DCF_0 from this point of view.) Here, I am interested in generalising this model-theoretic framework to *meromorphic* vector fields; namely, when X is a compact complex-analytic space that is not necessarily algebraic and v is a meromorphic section to the holomorphic tangent bundle. While DCF_0 is built on the theory of algebraically closed fields (ACF_0), the new theory I am seeking should be built on a first-order theory of compact complex manifolds.

R. Moosa was partially supported by an NSERC Discovery Grant.
MSC2020: primary 03C60; secondary 03C45, 32J99, 32M25.
Keywords: compact complex manifolds, meromorphic vector fields, differentially closed fields, model theory.

About thirty years ago, as part of the development of the notion of "Zariski-type structure", Zilber [1993] proposed a model theory for compact complex manifolds. Unlike ACF_0 and DCF_0, the first-order theory proposed by Zilber for compact complex manifolds was not given by an explicit axiomatisation, nor as the model companion of a natural class of algebraic structures, but rather as theories of particular structures: a compact complex manifold M is viewed as a first-order structure in the language where there is a predicate for each closed complex-analytic subset of each finite cartesian power of M. Zilber showed that the theory of any such structure shares many properties with its algebraic predecessors: in particular, they admit quantifier elimination and are of finite Morley rank (bounded by the dimension of M). Later, in [Hrushovski 1998; Pillay 2000], for example, it became common to consider all compact complex manifolds — indeed all (reduced) compact complex-analytic spaces — at once, in a multisorted structure whose theory now goes by the name CCM. Like differentially closed fields, CCM is a proper expansion of ACF_0. Also like DCF_0, much of the richness of geometric stability theory absent in ACF_0 is present in CCM. For example, all cases of the Zilber trichotomy appear.

In this paper, I present a common expansion of CCM and DCF_0, which I call DCCM. It turns out (in Section 8, below) that the finite-dimensional fragment of DCCM captures, precisely, the bimeromorphic geometry of meromorphic vector fields. As such, it achieves the goal set out in this introduction.

The theory DCCM arises by considering *differential* CCM-*structures*, essentially by adding a "derivation" to the definable closure of a generic point of a sort, say X, in CCM. This makes sense because the elements of the definable closure of a generic point of X can be viewed as meromorphic maps from X to other sorts, and hence can be differentiated. See Sections 2 and 3 for a detailed explanation.

The specific goals of this paper are:

(1) to show that the (universal) theory of differential CCM-structures admits a model companion, which is DCCM, by giving a geometric first-order axiomatisation of the existentially closed models (Theorem 5.5);

(2) to show that DCCM is complete, admits quantifier elimination (Proposition 6.3) and elimination of imaginaries (Theorem 7.6), and to give a geometric characterisation of definable and algebraic closure (Proposition 6.5);

(3) to show that DCCM is totally transcendental (Theorem 7.5), and to give geometric characterisations of nonforking independence (Corollary 7.4); and,

(4) to establish the correspondence between finite-dimensional types (over the empty set) in DCCM and meromorphic vector fields (Theorem 8.3).

The proofs proceed largely by finding geometric analogues for the algebraic arguments already familiar from DCF_0.

The next step in the study of DCCM, not attempted here, would be to establish the canonical base property for finite-dimensional types following the strategy of [Pillay and Ziegler 2003] in the case of DCF_0. This would involve developing a theory of jet spaces in DCCM; see for example [Bays et al. 2017], where this was done for compact complex manifolds with a generic automorphism (CCMA). In any case, once the canonical base property is established, a concrete manifestation of the Zilber dichotomy for finite-dimensional minimal types in DCCM will follow. It would then be reasonable to expect that many of the recent applications of model theory to algebraic vector fields, as carried out in [Freitag et al. 2022; Jaoui and Moosa 2022] for example, would extend to meromorphic vector fields.

The process of adding an automorphism to any given first-order theory of interest, and then seeking a model companion, is well-studied (see [Chatzidakis and Pillay 1998]). Here we have "added a derivation" instead. Clearly, this does not make sense for an arbitrary theory. But following the ideas presented here, it may be worth investigating a robust general setting where adding a derivation does make sense. A likely candidate might be that of Zariski-type structures in Zilber's sense; one that expands ACF_0 and admits a functor that extends to all sorts the tangent space construction on algebraic varieties.

2. Meromorphic varieties and their tangent spaces

In this section I want to slightly loosen the usual formalism for doing the model theory of compact complex-analytic spaces, so as to work directly in the "compactifiable" rather than compact setting.

For the fundamental notions from complex-analytic geometry we suggest [Fischer 1976]. Given a reduced compact complex-analytic space X, by the *Zariski topology* on X we mean the (noetherian) topology of closed complex-analytic subsets of X. This does not conflict with the usual meaning of the Zariski topology in the case that X is a projective complex-algebraic variety, because in that case the complex-analytic and complex-algebraic sets agree (Chow's theorem).

Definition 2.1. By a *meromorphic variety* we mean a pair $(X, \overline{X})$ where X is a Zariski dense and open subset of a reduced compact complex-analytic space $\overline{X}$. Note that X inherits from $\overline{X}$ the structure of a reduced complex-analytic space in its own right, which may admit other compactifications.[1] We usually abbreviate our notation by referring to X as the meromorphic variety, but it is important to keep in mind that we view X as embedded in a fixed given compactification $\overline{X}$.

Cartesian products of meromorphic varieties $X \times Y$ are viewed as meromorphic varieties with the compactification $\overline{X \times Y} = \overline{X} \times \overline{Y}$.

[1]Thanks to the anonymous referee for pointing out that different compactifications of X need not be bimeromorphically equivalent.

By the *Zariski topology* on X we mean the topology induced by the Zariski topology on $\overline{X}$. Note that this is a coarser topology than that of the closed complex-analytic subsets of X; such a set is Zariski closed in X if and only if its (euclidean) closure in $\overline{X}$ is Zariski closed.

By a *definable holomorphic map* $f : X \to Y$ of meromorphic varieties we mean a holomorphic map that extends to a meromorphic map $\bar{f} : \overline{X} \to \overline{Y}$. Equivalently, the graph of f is Zariski closed in $X \times Y$. More generally, a *definable meromorphic map* $f : X \to Y$ is a meromorphic map that extends to a meromorphic map from $\overline{X}$ to $\overline{Y}$. Such a map is *dominant* if its image is Zariski dense in Y.

Meromorphic varieties, as I have defined them here, are intended to extend the notion of quasiprojective variety from the complex-algebraic to the complex-analytic setting. Indeed, the quasiprojective varieties are precisely the meromorphic varieties X where $\overline{X}$ is projective algebraic. Note that while every regular or rational function on a quasiprojective variety X extends to a rational function on the projective closure $\overline{X}$, the same is not true of holomorphic and meromorphic functions on meromorphic varieties, and this is why we restrict our attention to *definable* holomorphic and meromorphic maps, namely, the ones that do so extend.

Remark 2.2. When $X = \overline{X}$ is compact, every holomorphic (respectively, meromorphic) map to a meromorphic variety, $f : X \to Y$, is definable holomorphic (respectively, definable meromorphic). This is because, by the proper mapping theorem, the image of f in Y is Zariski closed, and hence we can take $\bar{f}$ to be f itself, viewed as a map from X to $\overline{Y}$.

The usual model-theoretic set-up is to consider the first-order theory of the multisorted structure $\mathcal{A}$ where there is a sort for each reduced and irreducible compact complex-analytic space, and a predicate for each Zariski closed subset of each finite cartesian product of sorts. See, for example, the surveys [Moosa 2005a; Moosa and Pillay 2008]. Every meromorphic variety, in the above sense, is 0-definable in $\mathcal{A}$, as is every Zariski closed subset of every finite cartesian product of meromorphic varieties. It therefore does no harm to work instead with the expansion of $\mathcal{A}$ to the multisorted structure $\mathcal{M}$ where there is a sort for each irreducible meromorphic variety, and a predicate for each Zariski closed subset of each finite cartesian product of sorts. So we have added some sorts and some predicates, but they were all already 0-definable in the original structure. I denote by L the language of $\mathcal{M}$, and by CCM the first-order L-theory of $\mathcal{M}$. It admits quantifier elimination and elimination of imaginaries, and, sort by sort, is of finite Morley rank.

Every quasiprojective complex-algebraic variety V, given with an embedding in a projective compactification $\overline{V}$, is a meromorphic variety, and the algebraic and analytic Zariski topologies on V agree. In particular, definable holomorphic maps in this case are just regular morphisms, and definable meromorphic maps are rational. In this way, algebraic geometry lives as a pure reduct of CCM.

Our main use of the flexibility that $\mathcal{M}$ affords is that the collection of sorts is closed under taking tangent spaces. Recall that the *tangent space* of a complex-analytic space X is the linear fibre space $\pi : TX \to X$ associated to the sheaf of differentials $\underline{\Omega}_X^1$ on X. So TX is a complex-analytic space and $\pi : TX \to X$ is a surjective holomorphic map whose fibres are uniformly equipped with the structure of a complex vector space, in the sense that there are holomorphic maps for addition $+ : TX \times_X TX \to TX$, scalar multiplication $\lambda : \mathbb{C} \times TX \to TX$, and zero section $z : X \to TX$, all over X, satisfying the vector space axioms. For any point $p \in X$, the *tangent space to X at p* is the fibre of $\pi : TX \to X$ above p, denoted by T_pX, and it is canonically isomorphic as a complex vector space to $\operatorname{Hom}_{\mathbb{C}}(\mathfrak{m}_{X,p}/\mathfrak{m}_{X,p}^2, \mathbb{C})$, where $\mathfrak{m}_{X,p}$ is the maximal ideal of the local ring of X at p.

We claim that when X is a meromorphic variety so is TX, and that $\pi, +, \lambda, z$ are all definable holomorphic maps. Let us first consider the case when $X = \overline{X}$ is already compact. We are looking for a natural compactification of TX. In fact, there is a canonical way to do this for any linear fibre space $\mathcal{L}(\mathcal{F}) \to X$ associated to a coherent analytic sheaf $\mathcal{F}$ on X; it is just the relativisation of the usual embedding of $\mathbb{C}^n$ in the projectivisation of $\mathbb{C}^{n+1}$. One considers the coherent analytic sheaf $\mathcal{F} \times \mathcal{O}_X$ of rank one greater than $\mathcal{F}$, and then the associated projective linear space $\mathbb{P}(\mathcal{F} \times \mathcal{O}_X) \to X$. See [Fischer 1976, Section 1.9] for details. Then $\mathcal{L}(\mathcal{F})$ embeds in $\mathbb{P}(\mathcal{F} \times \mathcal{O}_X)$ over X as a Zariski open set in such a way that the linear structure (namely, $\pi, +, \lambda, z$) extends meromorphically to the projective linear space. Applying this to $\mathcal{F} = \underline{\Omega}_X^1$ gives $TX \to X$ the meromorphic structure we are looking for, namely

$$\overline{TX} := \mathbb{P}(\underline{\Omega}_X^1 \times \mathcal{O}_X).$$

Now, if we consider a general meromorphic variety X embedded in $\overline{X}$, then the linear space $TX \to X$ is just the restriction to X of $T\overline{X} \to \overline{X}$, and hence $\overline{T\overline{X}}$ serves as a compactification for TX, to which the linear structure extends meromorphically.

Remark 2.3. While we have not been assuming here that X is smooth, the tangent space is better behaved and more familiar under that assumption. As we are only interested in the bimeromorphic structure, we can achieve smoothness by replacing X with its nonsingular locus. Note that the set of nonsingular points of X is of the form $X \cap U$, where U is the (Zariski dense and open) set of nonsingular points of the compactification $\overline{X}$. It follows that the nonsingular locus of X is again a meromorphic variety given with the same compactification $\overline{X}$.

Recall that the tangent space construction is functorial: for each meromorphic (respectively, holomorphic) map $g : X \to Y$ between complex-analytic spaces there is a meromorphic (respectively, holomorphic) map $dg : TX \to TY$ such that

$$\begin{array}{ccc} TX & \xrightarrow{dg} & TY \\ {\scriptstyle \pi_X}\downarrow & & \downarrow{\scriptstyle \pi_Y} \\ X & \xrightarrow{g} & Y \end{array}$$

commutes, and we have the functoriality property $d(g \circ h) = (dg) \circ (dh)$. If X and Y are meromorphic varieties, and $g : X \to Y$ is definable meromorphic (respectively, holomorphic), then so is $dg : TX \to TY$. That is, if g extends to a meromorphic map $\overline{X} \to \overline{Y}$ then dg extends to a meromorphic map $\overline{TX} \to \overline{TY}$.

3. The differential structure

By a CCM-*structure* I mean a definably closed subset of a model of CCM. In other words, a model of $\mathrm{CCM}_\forall$. The goal of this section is to describe what we might consider a "derivation" on a CCM-structure. But first, let us recall what CCM-structures themselves look like.

Since we are in a relational language in which all elements of $\mathcal{M}$ are named, a model of $\mathrm{CCM}_\forall$ is simply a subset A of an elementary extension $\mathcal{N}$ of $\mathcal{M}$ such that $\mathcal{M} \subseteq A$. As we are in a multisorted setting, this is meant relative to every sort, so $S(\mathcal{M}) \subseteq S(A)$ for all sorts S of L. But we are mostly interested in finitely generated definably closed substructures, so where $A = \mathrm{dcl}(a)$ for some $a \in X(\mathcal{N})$ and some irreducible meromorphic variety X. Replacing X by the locus of a, we may assume that a is a generic point of X in the sense that it is not contained in $Y(\mathcal{N})$ for any proper Zariski closed subset $Y \subsetneq X$. In that case we can identify A with the set of all definable meromorphic maps $g : X \to S$ as S ranges over all other sorts. Indeed the identification is given by $g \mapsto g(a) \in S(A)$, noting that every point of $S(A)$ arises this way as $A = \mathrm{dcl}(a)$, and that if two definable meromorphic maps agree on a then they agree on X by genericity.

It is worth comparing to the algebraic case, so when X happens to be a quasi-projective complex-algebraic variety. In that case one only needs to consider the single target sort $S = \mathbb{P}$, the projective line. Indeed, in that case, $\mathrm{dcl}(a) = \mathbb{C}(X)$ is just the field of rational functions. For nonalgebraic meromorphic varieties, if we only considered $S = \mathbb{P}$ we would obtain the *meromorphic function field* of X, and not necessarily the full definable closure of a generic point. Indeed, on some compact complex-analytic spaces, namely those of *algebraic dimension* 0, there are no nonconstant meromorphic functions, but many nonconstant meromorphic maps to other sorts.

The differential structure I want to consider is motivated by the study of the following natural objects in bimeromorphic geometry:

Definition 3.1. By a *meromorphic vector field* we mean an irreducible meromorphic variety X equipped with a definable meromorphic section $v : X \to TX$ to the tangent space of X.

Remark 3.2. When $X = \overline{X}$ is compact, "definable" is redundant and a meromorphic vector field is simply a meromorphic section to the tangent space — see Remark 2.2. So this notion does generalise what I called a meromorphic vector field in the introduction. However, as we are only interested in the bimeromorphic geometry, it is not much of a generalisation: we can always pass from (X, v) to $(\overline{X}, \bar{v})$.

Of course, every meromorphic variety equipped with its zero section is a meromorphic vector field, which we call the *trivial vector field*.

Every (*rational*) *algebraic vector field*, by which we mean an irreducible quasi-projective complex-algebraic variety equipped with a rational section to the tangent space, is a meromorphic vector field. Indeed, these are the only meromorphic vector fields on algebraic varieties. In particular, as all compact complex-analytic spaces of dimension 1 are projective algebraic curves, every 1-dimensional meromorphic vector field is algebraic.

We already get nonalgebraic examples in dimension 2. As is pointed out in [Rebelo 2004, Example 2], for instance, all elliptic surfaces admit interesting meromorphic vector fields. Since there are nonalgebraic elliptic surfaces (every compact complex surface of algebraic dimension 1 is such), this is a class of nontrivial meromorphic vector fields that are not algebraic. These examples also show that meromorphic vector fields can be ubiquitous in situations where no holomorphic ones exist.

But there are also nonalgebraic holomorphic vector fields. Suppose $X = \overline{X}$ is compact and $G = \mathrm{Aut}_0(X)$ is the connected component of the automorphism group of X. Then G is a complex Lie group whose Lie algebra consists precisely of the holomorphic vector fields on X; see [Kobayashi 1972, Section III.1]. It follows that if $X = \overline{X}$ is nonalgebraic and $\mathrm{Aut}_0(X)$ is positive-dimensional, then X admits many nontrivial and nonalgebraic holomorphic (and hence meromorphic) vector fields. So, for example, if X is any complex torus, then $X = \mathrm{Aut}_0(X)$ acting by translation, and hence each point of the Lie algebra of X gives rise to an (invariant) holomorphic vector field on X.

Finally, it is worth noting, and was pointed out to me by the anonymous referee, that, unlike in the algebraic case, there are compact complex manifolds that admit no nontrivial meromorphic vector fields. For example, suppose X is a generic K3 surface. If X did admit a meromorphic vector field, v, then, as X has no proper infinite closed analytic subsets, the indeterminacy locus of v would be finite, and so, by Hartogs' theorem, v would extend to a holomorphic vector field on X. But K3 surfaces do not admit any nontrivial global holomorphic vector fields.

Suppose (X, v) is a meromorphic vector field, $\mathcal{N} \succeq \mathcal{M}$ is an elementary extension, and $a \in X(\mathcal{N})$ is a generic point of X. What structure does v induce on $A := \mathrm{dcl}(a)$? Well, for any definable meromorphic $g : X \to S$, we have the definable meromorphic map $\nabla_v(g) := dg \circ v : X \to TS$. Viewing $g \in S(A)$ we have defined a function $\nabla_v : S(A) \to TS(A)$, for all sorts S. Here are two salient properties of this function that are easily verified using the functoriality of the tangent space construction:

- $\pi \circ \nabla_v(g) = g$, where $\pi : TS \to S$ is the projection.
- $df \circ \nabla_v(g) = \nabla_v(f \circ g)$ for any definable meromorphic $f : S \to T$.

We are thus lead to consider the following notion:

Definition 3.3. Let $L_\nabla = L \cup \{\nabla\}$, where $\nabla = (\nabla_S : S \text{ sort of } L)$ and ∇_S is a function symbol from the sort S to the sort TS. Let $\mathrm{CCM}_{\forall,\nabla}$ denote the universal L_∇-theory which is obtained by adding to $\mathrm{CCM}_\forall$ the following axioms:

Axiom (1) For each sort S, $\nabla_S : S \to TS$ is a section to $\pi : TS \to S$.

Axiom (2) For each definable meromorphic map $f : S_1 \to S_2$ between sorts, the diagram

$$\begin{array}{ccc} TS_1 & \xrightarrow{df} & TS_2 \\ {\scriptstyle\nabla_{S_1}}\uparrow & & \uparrow{\scriptstyle\nabla_{S_2}} \\ S_1 & \xrightarrow{f} & S_2 \end{array}$$

commutes. Remembering that f and df are not function symbols in the language but rather their graphs are predicates, what we mean by this is the axiom

$$\forall xy\,\big((x, y) \in \Gamma(f) \Longrightarrow (\nabla_{S_1}x, \nabla_{S_2}y) \in \Gamma(df)\big).$$

We usually drop the subscript and write ∇ for ∇_S whenever it is clear from context which sort we are working in.

One consequence of Axiom (2) that gets used often without mention is that $\nabla(a_1, a_2) = (\nabla a_1, \nabla a_2)$ under the identification $T(S_1 \times S_2) = TS_1 \times TS_2$.

We can always extend uniquely to the definable closure:

Proposition 3.4. *Suppose $A \subseteq \mathcal{N} \models \mathrm{CCM}$ and $(A, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$. Then there is a unique extension of ∇ to $\mathrm{dcl}(A)$ making it a model of $\mathrm{CCM}_{\forall,\nabla}$.*

Proof. Let $B := \mathrm{dcl}(A)$. Given a sort S we need to define ∇ on $S(B)$. Fix $b \in S(B)$ and let $X := \mathrm{loc}(b) \subseteq S$ so that $b \in X(B)$ is generic. Since $b \in \mathrm{dcl}(A)$, there exists some other irreducible meromorphic variety Y admitting a dominant definable meromorphic map $f : Y \to X$, and a generic point $a \in Y(A)$, such that $b = f(a)$. Now, $df : TY \to TX$ and $\nabla(a) \in TY(A)$. Define $\nabla(b) := df(\nabla a)$. Indeed, this is forced upon us by Axiom (2) of Definition 3.3, and hence takes care of the uniqueness part of the statement.

We have to check that it is well-defined. Suppose we have another $f' : Y' \to X$ and $a' \in Y'(A)$ generic such that $b = f'(a')$ as well. Let $Z = \mathrm{loc}(a, a') \subseteq Y \times Y'$ and consider $\bar{f} := (f, f') : Z \to X^2$. Since $\bar{f}$ takes a generic point of Z to the diagonal $D \subseteq X^2$ we have that $\bar{f}(Z) \subseteq D$. Hence $d\bar{f} : TZ \to T(X^2)$ lands in TD, which is the diagonal in $T(X^2) = (TX)^2$. Since $d\bar{f}(\nabla(a, a')) = (df(\nabla a), df'(\nabla a'))$ by functoriality, this means that $df(\nabla a) = df'(\nabla a')$, as desired.

Next, observe that ∇ so defined is a function from $S(B)$ to $TS(B)$, and is a section to $\pi : TS \to S$. That is, (B, ∇) does satisfy Axiom (1) of Definition 3.3. Taking $f = \mathrm{id}$ in the above construction, we see also that $(A, \nabla) \subseteq (B, \nabla)$.

It remains to verify Axiom (2). That is, given $g : S_1 \to S_2$ a definable meromorphic map between sorts, and $b_i \in S_i(B)$ with $g(b_1) = b_2$, we need to show $dg(\nabla b_1) = \nabla b_2$. Note that by concatenating — namely, working in cartesian products — we can arrange things so that b_1 and b_2 are defined over the same tuple from A. That is, there is a sort S with $a \in S(A)$ such that $b_1 = f_1(a)$ and $b_2 = f_2(a)$, where $f_i : S \to S_i$ are definable meromorphic maps. Taking Zariski loci we may assume that a is generic in S and that each b_i is generic in S_i. Hence

$$\begin{aligned} dg(\nabla b_1) &= dg(df_1(\nabla a)) && \text{by how } \nabla \text{ is defined on } B \\ &= d(gf_1)(\nabla a) && \text{by functoriality} \\ &= df_2(\nabla a) && \text{as } gf_1 = f_2\text{, as that is the case on the generic } a \\ &= \nabla b_2 && \text{by how } \nabla \text{ is defined on } B, \end{aligned}$$

as desired. □

Definition 3.5. A *differential* CCM-*structure* is a model $(A, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$ such that $A = \mathrm{dcl}(A)$.

As a consequence of Proposition 3.4, when working with models of $\mathrm{CCM}_{\forall,\nabla}$ there is little loss of generality in assuming that we have a differential CCM-structure, namely that the underlying set is definably closed in CCM.

It is worth observing that standard points are always constant:

Lemma 3.6. *Suppose* (A, ∇) *is a differential* CCM-*structure and* S *is a sort. If* $p \in S(\mathcal{M})$ *then* $\nabla(p) = 0 \in T_pX$.

Proof. Note that $X := \{p\}$ is itself an irreducible meromorphic variety, and we can consider the containment as a definable holomorphic map $f : X \to S$. Now $TX = \{(p, 0)\}$, and hence $\nabla_X = 0$. But, by Axiom (2) of Definition 3.3, this forces

$$\nabla_S(p) = df(\nabla_X(p)) = 0$$

as $df_p : T_pX \to T_{f(p)}S$ is a linear map. □

In the finitely dcl-generated case we recover precisely the meromorphic vector fields that motivated Definition 3.3:

Proposition 3.7. *Suppose X is an irreducible meromorphic variety, a is a generic point of X in some elementary extension, and $A = \mathrm{dcl}(a)$. Then the differential* CCM*-structures on A are precisely the ∇_v induced by meromorphic vector fields $v : X \to TX$.*

Proof. We have already seen that $(A, \nabla_v) \models \mathrm{CCM}_{\forall,\nabla}$ if (X, v) is a meromorphic vector field. For the converse, suppose $(A, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$. Note that $a \in X(A)$ and $\nabla(a) \in TX(A)$. As definable meromorphic maps, $a \in X(A)$ is the identity map on X and $\nabla(a) \in TX(A)$ is some $v : X \to TX$. Axiom (1) ensures that v is a section to $\pi : TX \to X$, and hence a meromorphic vector field on X. It remains to verify that $\nabla = \nabla_v$. Let $g(a) \in S(A)$, where $g : X \to S$ is a definable meromorphic map and S is a sort. Then

$$\nabla_v(g(a)) = dg \circ v(a) = dg \circ \nabla(a) = \nabla(g(a)),$$

where the final equality is by Axiom (2). □

Note that Proposition 3.7 extends to meromorphic varieties the (well-known) correspondence, in the case when X is quasiprojective algebraic, between $\mathbb{C}$-linear derivations on $\mathbb{C}(X)$ and rational vector fields on X.

So the study of meromorphic vector fields amounts to the study of (finitely generated) differential CCM-structures. In the usual model-theoretic way, we will eventually look for a model companion: a theory that axiomatises the existentially closed differential CCM-structures.

We conclude this section by extending the notion of differential CCM-structure to a setting where ∇ is allowed to take values in an extension. This will be useful in what follows.

Definition 3.8. Suppose $\mathcal{N} \models \mathrm{CCM}$ and $A \subseteq \mathcal{N}$ is a definably closed set. By an *$\mathcal{N}$-valued differential* CCM*-structure* on A we mean a map $\nabla : S(A) \to TS(\mathcal{N})$, for every sort S, such that ∇ is a section to $\pi : TS \to S$, and such that $df(\nabla a) = \nabla(f(a))$ for all $a \in S(A)$ and all definable meromorphic maps f.

4. Prolongations

In this section we construct a version of the tangent space that is twisted by a differential structure. Since differential structure only has content in proper elementary extensions of $\mathcal{M}$, this will necessarily be about “meromorphic varieties over parameters” in arbitrary models of CCM, which we begin by reviewing.

Fix a model $\mathcal{N} \models \mathrm{CCM}$. Given an irreducible meromorphic variety X, we view it as a sort of L and consider its $\mathcal{N}$-points $X(\mathcal{N})$. Let us recall the Zariski topology on $X(\mathcal{N})$ with parameters from $\mathcal{N}$, sometimes referred to as the *nonstandard* Zariski topology to emphasise that we are not necessarily in the prime model $\mathcal{M}$.

See [Moosa 2004, Section 2] for a more detailed discussion. Every Zariski closed subset $Y \subseteq X(\mathcal{M})$ is named as a predicate in L and so we can consider $Y(\mathcal{N})$. These are the 0-definable Zariski closed subsets of $X(\mathcal{N})$. More generally, given a set of parameters $A \subseteq \mathcal{N}$, a Zariski closed subset $Y \subseteq X(\mathcal{N})$ over A is a subset of the form $Y = Z_a$ where $a \in S(A)$ is a generic point of another sort S and $Z \subseteq S \times X$ is a (0-definable) Zariski closed subset that projects dominantly on S. In diagrams:

$$\begin{tikzcd} Z \arrow[r, hook] \arrow[d, "\rho"'] & S \times X \\ S & \end{tikzcd}$$

That is, $Y = Z_a$ arises as the generic member of a 0-definable family of Zariski closed subsets of X. This forms a noetherian topology on $X(\mathcal{N})$. If Y is A-irreducible then we can take Z to be irreducible, and if Y is absolutely irreducible then we can take Z so that $\rho : Z \to S$ is a fibre space, meaning its general fibres in the standard model are irreducible.

The general standard fibres of $\rho : Z \to S$ are of constant dimension when Z is irreducible, giving rise to a notion of dimension for irreducible Zariski closed subsets of $X(\mathcal{N})$, which we denote by $\dim Y$.

The tangent space construction extends to nonstandard Zariski closed sets. Fix $Y = Z_a$ as above. Then the tangent spaces of the fibres of ρ in the standard model vary uniformly: Consider the diagram

$$\begin{tikzcd} S \times X & Z \arrow[l, hook'] \arrow[d, "\rho"'] & TZ \arrow[l] \arrow[r, hook] \arrow[d, "d\rho"] & TS \times TX \\ & S & TS \arrow[l] & \end{tikzcd}$$

and let $z : S \to TS$ be the zero section. For any $p \in S(\mathcal{M})$, the fibre $(TZ)_{z(p)} \subseteq TX$ of $d\rho$ above $z(p)$ is nothing other than $T(Z_p)$, the tangent space of $Z_p \subseteq X$. Hence we define the *tangent space of* $Y = Z_a$ in $\mathcal{N}$, denoted TY, to be $(TZ)_{z(a)}$.

Suppose, now, that $A = \operatorname{dcl}(A)$ and we have an $\mathcal{N}$-valued differential CCM-structure ∇ on A. Then, instead of considering the zero section, we can consider the differential section ∇. That is, since $\nabla(a) \in TS(\mathcal{N})$, we can consider the fibre $(TZ)_{\nabla(a)} \subseteq TX(\mathcal{N})$ of $d\rho$ over $\nabla(a)$. We define this to be the *prolongation space of* $Y = Z_a$, and denote it by τY. That is, $\tau Y := (TZ)_{\nabla(a)}$.

Lemma 4.1. *The above definition of τY depends only on Y and not on the presentation of Y as Z_a.*

Proof. Suppose Y also appears as Z'_b for some 0-definable Zariski closed $Z' \subseteq S' \times X$ with $b \in S'(A)$ generic. Replacing b with (a, b), we may assume that there is a dominant definable meromorphic map $f : S' \to S$ with $f(b) = a$, and that $Z' \subseteq Z \times_S S'$.

Hence $f' := (f, \mathrm{id}_X)\restriction_{Z'} : Z' \to Z$ restricts to the identity on $Z'_b = Y = Z_a$. Moreover, we have

$$\begin{tikzcd} Z & Z' \arrow[l, "f'"'] \arrow[d, "\rho'"] \\ S \arrow[u, leftarrow, "\rho"] & S' \arrow[l, "f"'] \end{tikzcd}$$

which yields

$$\begin{tikzcd} TZ \arrow[d, "d\rho"'] & TZ' \arrow[l, "df'"'] \arrow[d, "d\rho'"] \\ TS & TS' \arrow[l, "df"'] \end{tikzcd}$$

Since (A, ∇) is a differential CCM-structure, $df(\nabla b) = \nabla(a)$, so that

$$df'_{\nabla(b)} : (TZ')_{\nabla(b)} \to (TZ)_{\nabla(a)}.$$

Since $df' = (df, \mathrm{id}_{TX})\restriction_{TZ'}$, this shows that $(TZ')_{\nabla(b)} = (TZ)_{\nabla(a)}$, as desired. □

Remark 4.2. Given two such nonstandard Zariski closed sets Y_1, Y_2, there is a natural identification of $T(Y_1 \times Y_2)$ with $T(Y_1) \times T(Y_2)$ induced by the corresponding identification for (standard) meromorphic varieties. Moreover, if (A, ∇) is an ($\mathcal{N}$-valued) differential CCM-structure over which Y_1, Y_2 are defined, then we also have an identification of $\tau(Y_1 \times Y_2)$ with $\tau(Y_1) \times \tau(Y_2)$.

We denote the restriction of $\pi : TX \to X$ to τY also as $\pi : \tau Y \to Y$, and it is canonically attached to the prolongation space. For any $b \in Y$, we denote the fibre by $\tau_b Y$, and call it the *prolongation space to Y at b*. Note that if Y is an a-definable Zariski closed subset of X then τY is a $\nabla(a)$-definable Zariski closed subset of TX and $\tau_b Y$ is a $\nabla(a)b$-definable Zariski closed subset of $T_b X$.

Lemma 4.3. *Suppose that (B, ∇) is an $\mathcal{N}$-valued differential* CCM*-structure extending (A, ∇), and $b \in Y(B)$. Then $\nabla(b) \in \tau_b Y$.*

Proof. Since $b \in X(B)$ we must have $\nabla(b) \in T_b X$. So it remains to verify $\nabla(b) \in \tau Y$. Write $Y = Z_a$ as above. Then $(\nabla(a), \nabla(b)) = \nabla(a, b) \in TZ(\mathcal{N})$. In particular, $\nabla(b) \in (TZ)_{\nabla(a)}$, which is τY by construction. □

Lemma 4.4. *If Y is A-irreducible and $b \in Y$ is generic over A then $\tau_b Y$ is absolutely irreducible and* $\dim(\tau_b Y) = \dim Y$.

Proof. Let a from A be such that $Y = Z_a$ with $Z = \mathrm{loc}(a, b) \subseteq S \times X$ as above. Because $\rho : Z \to S$ is dominant, $d\rho$ restricts to a surjective $\mathbb{C}$-linear map between the tangent spaces at standard general points. Hence, at the generic point in $\mathcal{N}$, we have that $d\rho_{(a,b)} : T_{(a,b)} Z \to T_a S$ is a surjective $\mathbb{C}(\mathcal{N})$-linear map, where $\mathbb{C}(\mathcal{N})$ is the interpretation in $\mathcal{N}$ of the complex field, itself an algebraically closed field extending $\mathbb{C}$. By definition, the tangent space $T_b Y$ is the kernel of $d\rho_{(a,b)}$ while the prolongation space $\tau_b Y$ is $d\rho^{-1}_{(a,b)}(\nabla a)$. So $\tau_b Y$ is a coset of $T_b Y$ in $T_{(a,b)} Z$.

Absolute irreducibility of $\tau_b Y$ follows, and $\dim(\tau_b Y) = \dim(T_b Y)$. Finally, note that $\dim(T_b Y) = \dim Y$ because for standard general $(p, q) \in Z(\mathcal{M})$, the tangent space to Z_p at q is of dimension $\dim(Z_p)$. □

Finally, it is worth thinking about the case when Y is 0-definable, that is, using the above notation, when $a \in \mathcal{M}$. In that case, by Lemma 3.6, ∇ agrees with the zero section at a, and hence $\tau Y = TZ$ is just the tangent space of Y. That is, for 0-definable Zariski closed sets, the prolongation and tangent spaces agree.

5. Differentially closed CCM-structures

We aim to prove that $\mathrm{CCM}_{\forall,\nabla}$ admits a model companion. We begin by exploring some properties of the existentially closed (e.c.) models. This amounts to proving extension lemmas. For example, Proposition 3.4, which says that every model of $\mathrm{CCM}_{\forall,\nabla}$ extends to the definable closure of the underlying model of $\mathrm{CCM}_{\forall}$, implies that if (A, ∇) is an e.c. model of $\mathrm{CCM}_{\forall,\nabla}$ then it is a differential CCM-structure. Moreover, the e.c. models of $\mathrm{CCM}_{\forall,\nabla}$ are precisely the existentially closed differential CCM-structures. This justifies:

Definition 5.1. A *differentially closed* CCM-*structure* is an e.c. model of $\mathrm{CCM}_{\forall,\nabla}$.

Here is the main extension lemma:

Proposition 5.2. *Suppose* $\mathcal{N} \models \mathrm{CCM}$ *and* (A, ∇) *is an* $\mathcal{N}$*-valued differential* CCM-*structure. Suppose* X *is an irreducible meromorphic variety,* $b \in X(\mathcal{N})$*, and* $Y := \mathrm{loc}(b/A)$ *is the smallest* A*-definable Zariski closed subset of* $X(\mathcal{N})$*. For any* $c \in \tau_b Y$ *there is an extension of* ∇ *to an* $\mathcal{N}$*-valued differential* CCM-*structure on* $\mathrm{dcl}(Ab)$ *such that* $\nabla(b) = c$*.*

Proof. Let $D := \mathrm{dcl}(Ab)$. We follow the approach of Proposition 3.4. That is, given an element of D, say $d = f(a, b)$, where a is from A and $f : \mathrm{loc}(a, b) \to \mathrm{loc}(d)$ is a definable meromorphic map, we set $\nabla(d) := df(\nabla a, c)$. We have to verify that $(\nabla a, c) \in T_{(a,b)} \mathrm{loc}(a, b)$ for this to even make sense, that is, to be able to apply df to $(\nabla a, c)$. Note, first of all, that since $Y = \mathrm{loc}(b/A) \subseteq \mathrm{loc}(b)$ we do have that $c \in \tau_b Y \subseteq T_b \mathrm{loc}(b)$. So $(\nabla a, c) \in T_a \mathrm{loc}(a) \times T_b \mathrm{loc}(b)$, that is, $(\nabla a, c)$ lies above (a, b), and it only remains to check that $(\nabla a, c) \in T \mathrm{loc}(a, b)$. Since $Y = \mathrm{loc}(b/A) \subseteq \mathrm{loc}(b/a)$, and the latter is the fibre of the coordinate projection $\mathrm{loc}(a, b) \to \mathrm{loc}(a)$ over a, we have that $\tau Y \subseteq \tau \mathrm{loc}(b/a)$, and the latter is by definition the fibre of $T \mathrm{loc}(a, b) \to T \mathrm{loc}(a)$ over $\nabla(a)$. Since $c \in \tau Y$, this tells us that $(\nabla a, c) \in T \mathrm{loc}(a, b)$, as desired.

Considering the case when $d = a$ and $f : \mathrm{loc}(a, b) \to \mathrm{loc}(a)$ is the coordinate projection, we see that this definition of ∇ on D extends the given ∇ on A. Considering the case when $d = b$ (so that a is the empty tuple and $f = \mathrm{id}$), we see that $\nabla(b) = c$, as desired.

While the proof of Proposition 3.4 was carried out in the context of models of $\mathrm{CCM}_{\forall,\nabla}$, it works equally well in the setting of $\mathcal{N}$-valued differential CCM-structures, showing that the way we have defined ∇ on D above yields, for any sort S, a well-defined map $\nabla : S(D) \to TS(\mathcal{N})$ that is a section to $TS \to S$, and such that $dg(\nabla d) = \nabla(g(d))$ for any definable meromorphic map g and tuple $d \in S(D)$. So (D, ∇) is again an $\mathcal{N}$-valued differential CCM-structure. □

Corollary 5.3. *If* (A, ∇) *is a differentially closed* CCM*-structure then* $A \models \mathrm{CCM}$.

Proof. We have that $A \subseteq \mathcal{N}$ for some $\mathcal{N} \models \mathrm{CCM}$. Let (B, ∇) be a maximal $\mathcal{N}$-valued differential CCM-structure extending (A, ∇). This exists as $\mathcal{N}$-valued differential CCM-structures are preserved under unions of chains, as can be easily verified from the definition.

We claim that $B = \mathcal{N}$. Given $b \in X(\mathcal{N})$ for some sort X, let $Y = \mathrm{loc}(b/B)$ and choose $c \in \tau_b Y$. By Proposition 5.2 we can extend ∇ to an $\mathcal{N}$-valued differential CCM-structure on $\mathrm{dcl}(Bb)$. By maximality, it follows that $b \in X(B)$ to start with. As X and b were arbitrary, this shows that $B = \mathcal{N}$.

We have that $(A, \nabla) \subseteq (\mathcal{N}, \nabla)$ is an extension of differential CCM-structures. By quantifier elimination, CCM has a universal-existential axiomatisation. Since (A, ∇) is existentially closed, the truth of such axioms in $(\mathcal{N}, \nabla)$ implies their truth in (A, ∇). That is, $A \models \mathrm{CCM}$, as desired. □

This is, of course, not enough. That is, not every differential CCM-structure on a model of CCM is differentially closed. For example, the standard model $\mathcal{M}$ admits the trivial differential structure $\nabla = 0$, but is not existentially closed as we can use Proposition 5.2 to produce nontrivial differential CCM-structure extensions.

The following property of differentially closed CCM-structures, which we refer to as the *geometric axiom*, can be read as saying that ∇ is a "generic" section to the tangent space:

Proposition 5.4. *If* $(\mathcal{N}, \nabla)$ *is a differentially closed* CCM*-structure then it satisfies the following condition*:

(GA) *Suppose S is a sort, $X \subseteq S$ is an $\mathcal{N}$-definable irreducible Zariski closed subset, $Y \subseteq \tau X$ is an $\mathcal{N}$-definable irreducible Zariski closed subset that projects dominantly onto X, and $Y_0 \subsetneq Y$ is a proper $\mathcal{N}$-definable Zariski closed subset. Then there exists $a \in X(\mathcal{N})$ such that $\nabla(a) \in Y \setminus Y_0$.*

Proof. We already know, by Corollary 5.3, that $\mathcal{N} \models \mathrm{CCM}$. Let $\mathcal{U} \succeq \mathcal{N}$ be a sufficiently saturated elementary extension, and let $c \in Y(\mathcal{U})$ be generic in Y over $\mathcal{N}$. In particular, $c \in Y \setminus Y_0$. By dominance, $b := \pi(c) \in X(\mathcal{U})$ is generic over $\mathcal{N}$. In particular, $\mathrm{loc}(b/\mathcal{N}) = X$ and $c \in \tau_b X(\mathcal{U})$. So, by Proposition 5.2, we can extend ∇ to a $\mathcal{U}$-valued differential CCM-structure on $\mathrm{dcl}(\mathcal{N}b)$ such that $\nabla(b) = c$. Then, as in the proof of Corollary 5.3, we can extend ∇ further to all of $\mathcal{U}$ so that

$(\mathcal{U}, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$. Now, b witnesses that in $(\mathcal{U}, \nabla)$ there is a point of X that is sent by ∇ into $Y \setminus Y_0$. By existential closedness of $(\mathcal{N}, \nabla)$, there must exist $a \in X(\mathcal{N})$ such that $\nabla(a) \in Y \setminus Y_0$. □

As the terminology already indicates, the geometric axiom characterises differentially closed CCM-structures:

Theorem 5.5. *A model* $(\mathcal{N}, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$ *is existentially closed if and only if* $\mathcal{N} \models \mathrm{CCM}$ *and condition (GA) of Proposition 5.4 holds.*

Proof. Corollary 5.3 and Proposition 5.4 gave the left-to-right direction. We therefore assume that $\mathcal{N} \models \mathrm{CCM}$ and $(\mathcal{N}, \nabla)$ satisfies (GA), and show that $(\mathcal{N}, \nabla)$ is existentially closed. Let S be a sort, x a variable belonging to S, and $\phi(x)$ a (finite) conjunction of L_∇-literals over $\mathcal{N}$ that is realised by $c \in S(A)$ in some extension $(A, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$ of $(\mathcal{N}, \nabla)$. We need to show that $\phi(x)$ has a realisation already in $(\mathcal{N}, \nabla)$. As in the proof of Corollary 5.3, we can extend (A, ∇) further to $(\mathcal{U}, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$, where $\mathcal{U} \models \mathrm{CCM}$.

Let d be the *order* of $\phi(x)$, that is, the largest positive integer such that $\nabla^d(x)$, namely ∇ iterated d-times and applied to x, appears in $\phi(x)$. We leave it to the reader to verify that $\phi(x)$ can then be rewritten as $(\nabla^d(x) \in U) \wedge (\nabla^d(x) \notin V)$, where U and V are $\mathcal{N}$-definable Zariski closed subsets of $T^d(S)$, the d-th iterated tangent space of S.

Let $Y := \mathrm{loc}(\nabla^d c/\mathcal{N}) \subseteq T^d(S)(\mathcal{U})$. Since c realises $\phi(x)$, we must have that $\nabla^d(c) \in U \setminus V$ and so $Y \subseteq U$ and $Y \nsubseteq V$. In particular, $Y_0 := Y \cap V$ is a proper $\mathcal{N}$-definable Zariski closed subset of Y. We aim to find $a \in S(\mathcal{N})$ such that $\nabla^d(a) \in Y \setminus Y_0$; this suffices as such an a would be a realisation of $\phi(x)$ in $(\mathcal{N}, \nabla)$.

Let $\bar{c} := \nabla^{d-1}(c)$ and $X := \mathrm{loc}(\bar{c}/\mathcal{N})$. Then $\nabla^d(c) = \nabla(\bar{c})$, so that Y is contained in τX and projects dominantly onto X. Hence, by (GA), there is an $\bar{a} \in X(\mathcal{N})$ such that $\nabla(\bar{a}) \in Y \setminus Y_0$. Consider the first coordinate projection $\pi : T^{d-1}(S) \to S$, and set $a := \pi(\bar{a}) \in S(\mathcal{N})$. It suffices to show, therefore, that $\nabla^{d-1}(a) = \bar{a}$.

For each $\ell \geq 0$, let us denote by $\pi_\ell : T^{\ell+1}S \to T^\ell S$ the canonical projection. Moreover, for each $\ell = 0, \dots, d-1$, let $\bar{a}_\ell$ be the image of $\bar{a}$ in $T^\ell S$. So, in particular, $\bar{a}_0 = a$ and $\bar{a}_{d-1} = \bar{a}$. We claim that it suffices to show that

$$\bar{a}_{\ell+1} = \nabla(\bar{a}_\ell) \tag{5-1}$$

for all $\ell = 0, \dots, d-2$. Indeed, this would imply that

$$\bar{a} = \bar{a}_{d-1} = \nabla(\bar{a}_{d-2}) = \nabla^2(\bar{a}_{d-3}) = \cdots = \nabla^{d-1}(a),$$

as desired. So let us fix $\ell = 0, \dots, d-2$ and show (5-1). The idea is to construe (5-1) as a Zariski closed condition on $\nabla(\bar{a})$. First of all, noting that $\pi_{\ell+1}(\nabla\bar{a}_{\ell+1}) = \bar{a}_{\ell+1}$ and $d\pi_\ell(\nabla\bar{a}_{\ell+1}) = \nabla(\pi_\ell\bar{a}_{\ell+1}) = \nabla(\bar{a}_\ell)$, we see that (5-1) is equivalent to

$$\pi_{\ell+1}(\nabla\bar{a}_{\ell+1}) = d\pi_\ell(\nabla\bar{a}_{\ell+1}). \tag{5-2}$$

Next, letting $\rho : T^{d-1}S \to T^{\ell+1}S$ be the projection, we have that $\rho(\bar{a}) = \bar{a}_{\ell+1}$, and hence $\nabla(\bar{a}_{\ell+1}) = \nabla(\rho\bar{a}) = d\rho(\nabla(\bar{a}))$. So (5-2) is equivalent to

$$\pi_{\ell+1} d\rho(\nabla\bar{a}) = d(\pi_\ell\rho)(\nabla\bar{a}). \tag{5-3}$$

This is a Zariski closed condition on $\nabla\bar{a}$, and as $\nabla\bar{a}$ is in $Y = \operatorname{loc}(\nabla\bar{c}/\mathcal{N})$, it suffices to verify that the identity holds of $\nabla\bar{c}$. But this follows from the fact that $\nabla\bar{c} = \nabla^d c$,

$$\begin{aligned}
\pi_{\ell+1} d\rho(\nabla\bar{c}) &= \pi_{\ell+1} d\rho(\nabla^d c) \\
&= \pi_{\ell+1}\nabla(\rho(\nabla^{d-1}c)) \\
&= \rho(\nabla^{d-1}c) \\
&= \nabla^{\ell+1}c \\
&= \nabla(\nabla^\ell c) \\
&= \nabla(\pi_\ell\rho(\nabla^{d-1}c)) \\
&= d(\pi_\ell\rho)(\nabla^d c) \\
&= d(\pi_\ell\rho)(\nabla\bar{c}).
\end{aligned}$$

Hence (5-3) holds, as desired. $\square$

That condition (GA) of Proposition 5.4 is first-order expressible follows from the fact that as X varies in an L-definable family, τX varies in an L_∇-definable family by construction (see Section 4), and that in CCM irreducibility and domination are definable in parameters (see [Moosa 2004, Section 2]). Theorem 5.5 thus gives us a model companion to $\mathrm{CCM}_{\forall,\nabla}$, namely the theory of differentially closed CCM-structures, which we denote DCCM.

6. Basic model theory of DCCM

From general model theory, we have that DCCM is model-complete. In this section we prove that $\mathrm{CCM}_{\forall,\nabla}$ has the amalgamation property, from which we can deduce that DCCM is complete and admits quantifier elimination. As a consequence we obtain a geometric description of algebraic and definable closure.

But first we need an extension lemma for algebraic closure, whereas we have only dealt with definable closure (in Proposition 3.4) so far.

Lemma 6.1. *Suppose (A, ∇) is a differential* CCM-*structure with $A \subseteq \mathcal{N} \models$* CCM, *and $b \in \operatorname{acl}(A)$. Then there is a unique $\mathcal{N}$-valued differential* CCM-*structure on* $\operatorname{dcl}(Ab)$ *extending ∇. Moreover, this extension is in fact* $\operatorname{dcl}(Ab)$*-valued.*

Proof. Let $Y = \operatorname{loc}(b/A)$ and $c \in \tau_b Y$. By Proposition 5.2 we can extend ∇ from A to an $\mathcal{N}$-valued differential CCM-structure on $\operatorname{dcl}(Ab)$ by sending $\nabla(b) := c$. We show that $\tau_b Y = \{c\}$ is a singleton and hence $c \in \operatorname{dcl}(Ab)$, so that the above extension is in fact $\mathrm{dcl}(Ab)$-valued, and so $(\mathrm{dcl}(Ab), \nabla) \models \mathrm{CCM}_{\forall,\nabla}$. This also shows uniqueness as any extension of ∇ to $\mathrm{dcl}(Ab)$ would have to take b into $\tau_b Y = \{c\}$, by Lemma 4.3, and hence would agree with the one we just constructed.

Let $X := \mathrm{loc}(b)$ and write $Y = Z_a$, where $Z = \mathrm{loc}(a, b) \subseteq S \times X$ is a 0-definable irreducible Zariski closed set and S is a sort with $a \in S(A)$ generic. The fact that $b \in \mathrm{acl}(A)$ means that Y is finite, and hence the coordinate projection $\rho : Z \to S$ is generically finite-to-one. It follows that $d_p\rho : T_p Z \to T_{\rho(p)} S$ is an isomorphism for general $p \in Z(\mathcal{M})$. Hence $d_{(a,b)}\rho : T_{(a,b)} Z \to T_a S$ is a bijection. If $c, c' \in \tau_b Y$ then we know, by the proof of Proposition 5.2, that $(\nabla a, c), (\nabla a, c') \in T_{(a,b)} Z$. But as $d\rho$ takes both $(\nabla a, c)$ and $(\nabla a, c')$ to $\nabla(a) \in T_a S$ we must have $c = c'$. So $\tau_b Y = \{c\}$, as desired. □

Next we prove independent amalgamation. We use $\mathop{\smile\!\!\!\!\!\!|}^{\mathrm{CCM}}$ to mean nonforking independence in CCM.

Lemma 6.2. *Suppose A, B_1, B_2 are definably closed subsets of $\mathcal{N} \models$ CCM, with $A \subseteq B_1 \cap B_2$ and $B_1 \mathop{\smile\!\!\!\!\!\!|}_A^{\mathrm{CCM}} B_2$. Suppose ∇_i is a differential CCM-structure on B_i, for $i = 1, 2$, such that ∇_1 and ∇_2 agree on A. Then there is a common extension ∇ of ∇_1 and ∇_2 to $B := \mathrm{dcl}(B_1 B_2)$ such that $(B, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$.*

Proof. Using Lemma 6.1 we can extend the differential CCM-structure on A, B_1, B_2 uniquely to their algebraic closures in $\mathcal{N}$. In particular, ∇_1 and ∇_2 agree on $\mathrm{acl}(A)$. So we may as well assume that $A = \mathrm{acl}(A)$, and $B_i = \mathrm{acl}(B_i)$ for $i = 1, 2$. One consequence of A being algebraically closed is that Zariski loci over A are absolutely irreducible, and hence independence over A has the following Zariski-topological characterisation:

$$b_1 \mathop{\smile\!\!\!\!\!\!|}_A^{\mathrm{CCM}} b_2 \quad \text{if and only if} \quad \mathrm{loc}(b_1, b_2/A) = \mathrm{loc}(b_1/A) \times \mathrm{loc}(b_2/A).$$

See [Moosa 2004, Section 2] for details.

Every tuple from B is of the form $b = f(b_1, b_2)$, where each b_i is from B_i, and $f : \mathrm{loc}(b_1, b_2) \to \mathrm{loc}(b)$ is a definable meromorphic map. Our only choice is to define

$$\nabla(b) := df(\nabla_1 b_1, \nabla_2 b_2).$$

But we need $(\nabla_1 b_1, \nabla_2 b_2) \in T_{(b_1,b_2)} \mathrm{loc}(b_1, b_2)$ for this to make sense. This is what we now check.

Let a be a tuple from A such that $\mathrm{loc}(b_1, b_2/A) = \mathrm{loc}(b_1, b_2/a)$. Let us denote by ∇ the common restriction of ∇_1 and ∇_2 to A. Taking prolongations with respect to the differential CCM-structure (A, ∇), and using the fact that for $i = 1, 2$ we have $(A, \nabla) \subseteq (B_i, \nabla_i)$, we see that $\nabla_i(b_i) \in \tau_{b_i} \operatorname{loc}(b_i/a)$. Hence,

$$\begin{aligned}(\nabla_1 b_1, \nabla_2 b_2) &\in \tau_{b_1} \operatorname{loc}(b_1/a) \times \tau_{b_2} \operatorname{loc}(b_2/a) \\ &= \tau_{(b_1,b_2)}(\operatorname{loc}(b_1/a) \times \operatorname{loc}(b_2/a)) \\ &= \tau_{(b_1,b_2)} \operatorname{loc}(b_1, b_2/a) && \text{as } b_1 \underset{a}{\overset{\mathrm{CCM}}{\perp}} b_2 \\ &\subseteq T_{(b_1,b_2)} \operatorname{loc}(b_1, b_2)\end{aligned}$$

as desired.

So it does make sense to set $\nabla(b) := df(\nabla_1 b_1, \nabla_2 b_2)$ for $b = f(b_1, b_2)$. The next step is to make sure this is well-defined. What if we also have $b = f'(b_1', b_2')$? This is dealt with exactly as in Proposition 3.4. Namely, let $Z := \operatorname{loc}(b_1, b_2, b_1', b_2')$ and consider $\bar{f} := (f, f') : Z \to \operatorname{loc}(b)^2$. Since $\bar{f}$ takes a generic point of Z to the diagonal we have that $d\bar{f} : TZ \to T(\operatorname{loc}(b)^2) = (T \operatorname{loc}(b))^2$ lands in the diagonal. Now, the argument in the previous paragraph, applied to $b_i b_i'$, shows, in particular, that $(\nabla_1(b_1 b_1'), \nabla_2(b_2 b_2')) \in T \operatorname{loc}(b_1 b_1', b_2 b_2')$. Hence, $(\nabla_1 b_1, \nabla_2 b_2, \nabla_1 b_1', \nabla_2 b_2') \in TZ$ and we get that $df(\nabla_1 b_1, \nabla_2 b_2) = df'(\nabla_1 b_1', \nabla_2 b_2')$.

We have defined ∇ on B in such a way that it is a section to $TS \to S$ for any sort S. It remains to check Axiom (2) of Definition 3.3. That is, suppose $g : S \to S'$ is a definable meromorphic map between sorts, and $b \in S(B)$, $b' \in S'(B)$ with $g(b) = b'$. We need to show that $dg(\nabla b) = \nabla b'$. We may assume that there are b_1, b_2 from B_1, B_2, respectively, and definable meromorphic maps f, f' such that $b = f(b_1, b_2)$ and $b' = f'(b_1, b_2)$. It follows that $gf = f'$ on $\operatorname{loc}(b_1, b_2)$, and so we compute

$$\begin{aligned}dg(\nabla b) &= dg(df(\nabla_1 b_1, \nabla_2 b_2)) && \text{by definition of } \nabla(b) \\ &= d(gf)(\nabla_1 b_1, \nabla_2 b_2) && \text{by functoriality} \\ &= df'(\nabla_1 b_1, \nabla_2 b_2) && \text{as } gf = f' \\ &= \nabla b' && \text{by definition of } \nabla(b').\end{aligned}$$

This completes the proof that $(B, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$. □

Proposition 6.3. $\mathrm{CCM}_{\forall,\nabla}$ *has the amalgamation property. In particular*, DCCM *admits quantifier elimination and is complete.*

Proof. Suppose $(B_i, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$, for $i = 1, 2$, with a common substructure (A, ∇). We seek a model $(B, \nabla) \models \mathrm{CCM}_{\forall,\nabla}$ into which (B_1, ∇) and (B_2, ∇) both embed over A. Let $\mathcal{U} \supseteq B_1$ be a sufficiently saturated model of CCM. By universality, there is an embedding of B_2 into $\mathcal{U}$ over A. Moreover, after taking nonforking extensions in CCM, we can find such an embedding such that the image of B_2 is independent from B_1 over A. We may as well assume, therefore, that $B_2 \subseteq \mathcal{U}$ already, and that $B_1 \perp_A^{\mathrm{CCM}} B_2$. Applying Lemma 6.2, we have a differential CCM-structure ∇ on $B := \operatorname{dcl}(B_1 B_2)$ that extends ∇ on both B_1 and B_2.

Quantifier elimination now follows for DCCM, as a general consequence for a model companion of a universal theory with amalgamation.

Completeness also follows as we have a prime substructure: all differentially closed CCM-structures extend the standard model $\mathcal{M} \models \mathrm{CCM}$ equipped with the trivial differential structure (Lemma 3.6). $\square$

Remark 6.4. In the case of DCF_0 quantifier elimination implies that every definable set is a finite boolean combination of the closed sets of a certain noetherian topology, namely the Kolchin topology. There is a natural analogue of the Kolchin topology here, a *meromorphic Kolchin topology* on each sort, which one expects to also be noetherian. I leave this to the interested reader to pursue.

Next, we wish to characterise definable and algebraic closure in DCCM. First of all, given $(\mathcal{N}, \nabla) \models \mathrm{DCCM}$ and $A \subseteq \mathcal{N}$, let us denote by $\langle A \rangle$ the L_∇-structure generated by A. If A is already an L_∇-substructure and a is a tuple then we denote by $A\langle a \rangle$ the L_∇-structure generated by $A \cup \{a\}$. Note that

$$A\langle a \rangle = A \cup \{a, \nabla(a), \nabla^2(a), \dots\}.$$

Quantifier elimination tells us that $\mathrm{tp}(a/A) = \mathrm{tp}(a'/A)$ if and only if there is an L-isomorphism $\alpha : A\langle a \rangle \to A\langle a' \rangle$ that fixes A pointwise and sends $\nabla^n(a)$ to $\nabla^n(a')$ for all $n \geq 0$.

We have been using acl and dcl for algebraic and definable closure in the L-theory CCM. We continue to do so, using acl_∇ and dcl_∇ for algebraic and definable closure in the L_∇-theory DCCM.

Proposition 6.5. *Suppose* $(\mathcal{N}, \nabla)$ *is a differentially closed* CCM-*structure and* $A \subseteq \mathcal{N}$. *Then* $\mathrm{dcl}_\nabla(A) = \mathrm{dcl}(\langle A \rangle)$ *and* $\mathrm{acl}_\nabla(A) = \mathrm{acl}(\langle A \rangle)$.

Proof. By Proposition 3.4, $\mathrm{dcl}(\langle A \rangle)$ is a differential CCM-substructure of $(\mathcal{N}, \nabla)$. Replacing A by $\mathrm{dcl}(\langle A \rangle)$, we may as well assume that A is a differential CCM-substructure to start with, and show that $\mathrm{dcl}_\nabla(A) = A$ and $\mathrm{acl}_\nabla(A) = \mathrm{acl}(A)$. The right-to-left containments are clear.

For the converses, let us first suppose that $b \notin \mathrm{acl}(A) =: B$. By Lemma 6.1, (B, ∇) is a differential CCM-substructure of $(\mathcal{N}, \nabla)$. We can find, in some elementary extension $\mathcal{U}$ of $\mathcal{N}$, a copy of $\mathcal{N}$ over B, say $\mathcal{N}'$, such that $\mathcal{N} \mathop{\smile\!\!\!\!\!\mid}_B^{\mathrm{CCM}} \mathcal{N}'$. Let $\alpha : \mathcal{N} \to \mathcal{N}'$ be an L-isomorphism over B witnessing this, and consider $b' := \alpha(b')$. The fact that $b \mathop{\smile\!\!\!\!\!\mid}_B^{\mathrm{CCM}} b'$ and that $b \notin B$ forces $b \neq b'$. On the other hand, setting $\nabla' := \alpha \nabla \alpha^{-1}$ we have that $(\mathcal{N}', \nabla') \models \mathrm{DCCM}$ and that $\alpha : (\mathcal{N}, \nabla) \to (\mathcal{N}', \nabla')$ is an L_∇-isomorphism over B. Now, we can find a common extension of ∇ and ∇' to $\mathrm{dcl}(\mathcal{N}\mathcal{N}')$ in $\mathcal{N}$ by Lemma 6.2 and then further to a model $(\mathcal{K}, \nabla) \models \mathrm{DCCM}$. So, in $(\mathcal{K}, \nabla)$ we have produced at least two distinct realisations, b and b', of $\mathrm{tp}(b/B)$. Repeating the process we can show that $\mathrm{tp}(b/B)$ has arbitrarily many realisations. That is, $b \notin \mathrm{acl}_\nabla(B) = \mathrm{acl}_\nabla(A)$, as desired.

Finally, suppose, toward a contradiction, that $b \in \mathrm{dcl}_\nabla(A) \setminus A$. This time we produce two distinct realisations of $\mathrm{tp}(b/A)$ for our contradiction. Since $\mathrm{dcl}_\nabla(A) \subseteq \mathrm{acl}_\nabla(A) = \mathrm{acl}(A)$, we have that $b \in \mathrm{acl}(A) \setminus A$. Hence $\mathrm{tp}_L(b/A)$ has a second realisation, $b' \in \mathrm{acl}(A)$ with $b' \neq b$. We thus have an L-isomorphism $\alpha : \mathrm{dcl}(Ab) \to \mathrm{dcl}(Ab')$ that fixes A pointwise and sends b to b'. But, by Lemma 6.1, $\mathrm{dcl}(Ab)$ and $\mathrm{dcl}(Ab')$ are differential CCM-substructures of $(\mathcal{N}, \nabla)$, and, as they each admit unique differential structures extending ∇ on A, we must have that α is an L_∇-isomorphism. By quantifier elimination, this means $\mathrm{tp}(b/A) = \mathrm{tp}(b'/A)$. $\square$

7. Stability and elimination of imaginaries

We work now in a fixed sufficiently saturated model $(\mathcal{U}, \nabla) \models \mathrm{DCCM}$ and adopt the usual convention that all parameter sets are assumed to be of cardinality less than that of the saturation.

In order to prove that DCCM is a stable theory, and to capture the meaning of nonforking independence therein, we follow an axiomatic approach. That is, we first introduce a natural notion of independence and then show that it has all the properties that characterise nonforking independence in stable theories.

Definition 7.1. Given sets A, B, C, we say that *A is independent of B over C*, denoted by $A \mathop{\smile\hskip-0.9em\vert}_C B$, to mean that $\mathrm{dcl}_\nabla(A) \mathop{\smile\hskip-0.9em\vert}^{\mathrm{CCM}}_{\mathrm{dcl}_\nabla(C)} \mathrm{dcl}_\nabla(B)$.

Note that we do not yet know that $\mathop{\smile\hskip-0.9em\vert}$ is nonforking independence, but we allow ourselves the notation as we will soon see that it is.

Let us first verify that $\mathop{\smile\hskip-0.9em\vert}$ is a notion of independence, in the sense introduced in [Kim and Pillay 1997]. First of all, it is clearly invariant under the action of automorphisms of $(\mathcal{U}, \nabla)$. Local character, finite character, symmetry, and transitivity all follow easily from the corresponding properties for $\mathop{\smile\hskip-0.9em\vert}^{\mathrm{CCM}}$.

Lemma 7.2 (extension). *Given $a, C \subseteq B$ there is $a' \models \mathrm{tp}(a/C)$ such that $a' \mathop{\smile\hskip-0.9em\vert}_C B$.*

Proof. We may assume that $C = \mathrm{dcl}_\nabla(C)$ and $B = \mathrm{dcl}_\nabla(B)$. By extension in CCM there is a sequence $(a'_n : n \geq 0) \mathop{\smile\hskip-0.9em\vert}^{\mathrm{CCM}}_C B$ and an L-isomorphism

$$\alpha : C\langle a\rangle \to C \cup \{a'_n : n \geq 0\}$$

that fixes C pointwise and takes $\nabla^n(a)$ to a'_n for all $n \geq 0$. Extend α to an L-isomorphism

$$\alpha : A := \mathrm{dcl}(C\langle a\rangle) \to A' := \mathrm{dcl}(C \cup \{a'_n : n \geq 0\}).$$

Set $\nabla' := \alpha\nabla\alpha^{-1}$ on A' so that (A', ∇') is a differential CCM-structure isomorphic to (A, ∇). On the other hand, since $A' \mathop{\smile\hskip-0.9em\vert}^{\mathrm{CCM}}_C B$, Lemma 6.2 gives us a common extension of (A', ∇') and (B, ∇) to a model $(\mathcal{N}, \nabla') \models \mathrm{DCCM}$. By universality we have an embedding $\iota : (\mathcal{N}, \nabla') \to (\mathcal{U}, \nabla)$ over B. Then $\beta := \iota \circ \alpha : A \to \mathcal{U}$ is an L-isomorphism with its image that fixes C pointwise and takes $\nabla^n(a)$ to $\nabla^n(\beta(a))$ for all $n \geq 0$. Hence, by quantifier elimination, $a' := \beta(a) \models \mathrm{tp}(a/C)$. Now $A' \mathop{\smile\!\!\!\!\!\!\mid}_C^{\mathrm{CCM}} B$ implies that $\iota(A') \mathop{\smile\!\!\!\!\!\!\mid}_C^{\mathrm{CCM}} B$ as ι is over B. But $\iota(A') = \beta(A) = \mathrm{dcl}(C\langle a'\rangle)$. So $a' \mathop{\smile\!\!\!\!\!\!\mid}_C B$, as desired. □

Lemma 7.3 (stationarity over algebraically closed sets). *Suppose* $C = \mathrm{acl}_\nabla(C) \subseteq B$, *and* a, a' *are tuples. If* $\mathrm{tp}(a/C) = \mathrm{tp}(a'/C)$, *and both* a *and* a' *are independent of* B *over* C, *then* $\mathrm{tp}(a/B) = \mathrm{tp}(a'/B)$.

Proof. We may assume, without loss of generality, that $B = \mathrm{dcl}_\nabla(B)$. Since $\mathrm{tp}(a/C) = \mathrm{tp}(a'/C)$, there is an L-isomorphism $\alpha : C\langle a\rangle \to C\langle a'\rangle$ that fixes C pointwise and takes $\nabla^n(a)$ to $\nabla^n(a')$ for all $n \geq 0$. Since C is algebraically closed, and the sequences $(\nabla^n a : n \geq 0)$ and $(\nabla^n a' : n \geq 0)$ are both CCM-independent from B over C, stationarity over algebraically closed sets in CCM implies that there is an L-isomorphism $\beta : B\langle a\rangle \to B\langle a'\rangle$ that fixes B pointwise and takes $\nabla^n(a)$ to $\nabla^n(a')$ for all $n \geq 0$. By quantifier elimination, $\mathrm{tp}(a/B) = \mathrm{tp}(a'/B)$. □

Corollary 7.4. DCCM *is a stable theory and* $\mathop{\smile\!\!\!\!\!\!\mid}$ *is nonforking independence.*

Proof. This follows from the above observations by the characterisation of nonforking independence in simple (and hence stable) theories as in [Kim and Pillay 1997]. □

We can also deduce stability by counting types. In fact, we get total transcendentality:

Theorem 7.5. DCCM *is* λ*-stable for every cardinal* $\lambda \geq 2^{\aleph_0}$.

Proof. We count types. Fix $\lambda \geq 2^{\aleph_0}$ and a subset $A \subseteq \mathcal{U}$ of cardinality at most λ. We show that there are at most λ-many complete types over A. We may assume that $A = \mathrm{dcl}_\nabla(A)$ is a differential CCM-substructure.

Suppose X is a sort, $a \in X(\mathcal{U})$, and consider $\mathrm{tp}(a/A)$. By quantifier elimination, it is determined by the sequence of types $(\mathrm{tp}_L(\nabla^n a/A) : n \geq 0)$ in CCM. Let

$$Z_n := \mathrm{loc}(\nabla^n a/A\nabla^{n-1}a).$$

I claim that there is some $N \geq 0$ such that $Z_{n+1} = \tau_{\nabla^n a}(Z_n)$, for all $n \geq N$. This suffices, since then $\mathrm{tp}(a/A)$ is determined by the pair $(N, \mathrm{tp}_L(\nabla^N a/A))$, of which there are at most λ-many possibilities by the λ-stability of CCM.

Note that $\nabla^{n+1}(a) \in \tau_{\nabla^n a}(Z_n)$, and so $Z_{n+1} \subseteq \tau_{\nabla^n a}(Z_n)$. But $\dim(\tau_{\nabla^n a}(Z_n)) = \dim(Z_n)$ by Lemma 4.4. So $\dim(Z_{n+1})$ is a nonincreasing function of n that must eventually stabilise. By irreducibility of $\tau_{\nabla^n a}(Z_n)$, this forces $Z_{n+1} = \tau_{\nabla^n a}(Z_n)$ for large enough n, as desired. □

Theorem 7.6. DCCM *admits elimination of imaginaries.*

Proof. A general criterion for elimination of imaginaries in a stable theory is that finite sets have codes and global types have canonical bases in the home sorts; see, for example, [Johnson 2020, Section 3]. That finite sets in DCCM have codes in the home sort follows from elimination of imaginaries in CCM; see [Pillay 2000; Moosa 2005b, Appendix].

So, we fix a saturated $\mathcal{N} \preceq \mathcal{U}$ and a complete type $p = \mathrm{tp}(a/\mathcal{N})$, and show that p has a canonical base in $\mathcal{N}$. Let N be as in the proof of Theorem 7.5; that is,

$$\mathrm{loc}(\nabla^{n+1}a/\mathcal{N}\nabla^n a) = \tau_{\nabla^n a}(\mathrm{loc}(\nabla^n a/\mathcal{N}\nabla^{n-1}a)),$$

for all $n \geq N$. By elimination of imaginaries for CCM, there is a code c for $\mathrm{loc}(\nabla^N a/\mathcal{N})$ in $\mathcal{N}$. We claim that c is a canonical base for p.

Fix an L_∇-automorphism σ of $\mathcal{N}$. We need to show that $p^\sigma = p$ if and only if $\sigma(c) = c$. One direction is clear: if $p^\sigma = p$ then $\mathrm{tp}_L(\nabla^N a/\mathcal{N})^\sigma = \mathrm{tp}_L(\nabla^N a/\mathcal{N})$ and hence $\mathrm{loc}(\nabla^N a/\mathcal{N})^\sigma = \mathrm{loc}(\nabla^N a/\mathcal{N})$, so that $\sigma(c) = c$.

For the converse, suppose $\sigma(c) = c$. Extend σ to an L_∇-automorphism $\hat{\sigma}$ of $\mathcal{U}$, and let $\hat{a} := \hat{\sigma}(a)$. We have that

$$\begin{aligned}
\mathrm{tp}_L(\nabla^N \hat{a}/\mathcal{N}) &= \mathrm{tp}_L(\nabla^N a/\mathcal{N}) && \text{since } \sigma(c) = c,\\
\mathrm{loc}(\nabla^{N+1}a/\mathcal{N}\nabla^N a) &= \tau_{\nabla^N a}(\mathrm{loc}(\nabla^N a/\mathcal{N}\nabla^{N-1}a)) && \text{by choice of } N\text{, and}\\
\mathrm{loc}(\nabla^{N+1}\hat{a}/\mathcal{N}\nabla^N \hat{a}) &= \tau_{\nabla^N \hat{a}}(\mathrm{loc}(\nabla^N \hat{a}/\mathcal{N}\nabla^{N-1}\hat{a})) && \text{by applying } \hat{\sigma}.
\end{aligned}$$

These imply that $\mathrm{tp}_L(\nabla^{N+1}\hat{a}/\mathcal{N}) = \mathrm{tp}_L(\nabla^{N+1}a/\mathcal{N})$. We can iterate to prove that $\mathrm{tp}_L(\nabla^n \hat{a}/\mathcal{N}) = \mathrm{tp}_L(\nabla^n a/\mathcal{N})$ for all $n \geq 0$. By quantifier elimination,

$$p^\sigma = \mathrm{tp}(\hat{a}/\mathcal{N}) = \mathrm{tp}(a/\mathcal{N}) = p,$$

as desired. □

8. Meromorphic vector fields and finite-dimensional types

We return in this final section to the motivating objects of interest: meromorphic vector fields. Our goal is to show that they are captured, up to bimeromorphic equivalence, in DCCM by the "finite-dimensional" types.

We continue to work in a fixed sufficiently saturated model $(\mathcal{U}, \nabla) \models \mathrm{DCCM}$.

Definition 8.1. Suppose A is an L_∇-substructure and $p = \mathrm{tp}(b/A)$ is a complete type. By the *dimension of* p, denoted by $\dim_\nabla(p)$ or $\dim_\nabla(b/A)$, we mean the sequence of nonnegative nondecreasing integers $\big(\dim(\mathrm{loc}(\nabla^n b/A)) : n \geq 0\big)$ ordered lexicographically. If $\dim_\nabla(p)$ is eventually constant then we say that p is *finite-dimensional* and we (re)use $\dim_\nabla(p)$ to denote that eventual finite number.

Note that the dimension depends only on the type p and not on the choice of realisation b. On the other hand, this dimension is not invariant under definable bijection — for example, b and $\nabla(b)$ are interdefinable over the empty set but the dimension sequences are not always the same (one is a shift of the other).

Nevertheless, whether or not a type is finite-dimensional, and the value of that finite dimension in the case that it is, is invariant under definable bijection.

Dimension witnesses forking:

Proposition 8.2. *Suppose a is a tuple and $C \subseteq B$ are L_∇-substructures. Then $a \mathop{\smile\!\!\!\!\!\mid}_C B$ if and only if* $\dim_\nabla(a/B) = \dim_\nabla(a/C)$.

Proof. We may assume that $B = \mathrm{dcl}_\nabla(B)$ and $C = \mathrm{dcl}_\nabla(C)$. By Proposition 6.5, we have that $\mathrm{dcl}_\nabla(Ca) = \mathrm{dcl}(C\langle a\rangle)$. Hence $a \mathop{\smile\!\!\!\!\!\mid}_C B$ is equivalent to $\nabla^n(a) \mathop{\smile\!\!\!\!\!\mid}_C^{\mathrm{CCM}} B$ for all $n \geq 0$. But, as dimension witnesses forking in CCM, this is equivalent to $\dim(\mathrm{loc}(\nabla^n a/B)) = \dim(\mathrm{loc}(\nabla^n a/C))$ for all $n \geq 0$. □

It follows that if p is finite-dimensional then it is of finite U-rank, and in fact that $U(p) \leq \dim_\nabla(p)$. One can ask whether the same holds of Morley rank: is it the case that Morley rank is bounded by dimension? It is also natural to ask about the converse: if p is of finite rank (U-rank or Morley rank, it is the same thing) must it be of finite dimension? One expects affirmative answers to these questions, as is the case for DCF_0, but I do not pursue them here.

A natural source of finite-dimensional types over the empty set are meromorphic vector fields in the sense of Definition 3.1. Suppose (X, v) is such. Consider the type $p(x)$, over the empty set, which says that $x \in X$ is generic and that $\nabla(x) = v(x)$. This is consistent by the geometric axiom of Proposition 5.4. Indeed, given any proper Zariski closed $X_0 \subseteq X$, apply (GA) to Y the Zariski closure of the image of v in $TX = \tau X$ and Y_0 the restriction of Y to X_0, yielding a $\mathcal{U}$-point $a \in X \setminus X_0$ with $\nabla(a) = v(a)$. Moreover, by quantifier elimination, this type is complete: the L-type of x is determined by x being generic in X, and $\nabla(x) = v(x)$ implies $\nabla^n(x) = v_n(x)$ for appropriate definable meromorphic $v_n : X \to T^n X$, for all n. We call p the *generic type of* (X, v). Note that p is finite-dimensional; in fact, $\dim_\nabla(p) = \dim X$. Indeed, for all $n \geq 0$, we have that $\nabla^n(b) = v_n(b)$ and v_n is a definable meromorphic section to $T^n X \to X$, and hence, as b is generic in X, we get $\dim(\mathrm{loc}(\nabla^n b/A)) = \dim X$.

It turns out that all finite-dimensional types arise this way:

Theorem 8.3. *Every finite-dimensional type over the empty set in* DCCM *is, up to interdefinability, the generic type of a meromorphic vector field.*

Proof. Suppose $p = \mathrm{tp}(b)$ is finite-dimensional. Let $d \geq 0$ be such that

$$\dim(\mathrm{loc}(\nabla^{d+1} b)) = \dim(\mathrm{loc}(\nabla^d b)).$$

Since the projection $\mathrm{loc}(\nabla^{d+1} b) \to \mathrm{loc}(\nabla^d b)$ is dominant, this means that it must be generically finite-to-one. Hence, setting $c := \nabla^{d+1}(b)$, we have that $c \in \mathrm{acl}(\nabla^d(b))$. As in the proof of Lemma 6.1, it follows that if $Y = \mathrm{loc}(c/\nabla^d(b))$ then $\tau_c Y$ is a singleton. Since Y is defined over $\nabla^d(b)$, the prolongation space τY is defined over $\nabla^{d+1}(b) = c$, and hence also $\tau_c Y$ is defined over c. By Lemma 4.3, $\nabla(c) \in \tau_c Y$. So $\nabla(c) \in \operatorname{dcl}(c)$. By quantifier elimination in CCM we can write $\nabla(c) = v(c)$ for some definable meromorphic map v. Then, setting $X := \operatorname{loc}(c)$, we have that $v : X \to TX$ is a section to the tangent space of X. That is, (X, v) is a meromorphic vector field and $q = \operatorname{tp}(c)$ is its generic type. Finally, observe that b and $c = \nabla^{d+1}(b)$ are interdefinable over the empty set, so that p and q are interdefinable types. □

The upshot is that the finite-dimensional fragment of DCCM, over the empty set, captures precisely the bimeromorphic geometry of meromorphic vector fields.

Remark 8.4. We have restricted our attention in this discussion to the empty set for brevity; we could have worked more generally over arbitrary parameters A. The result would be that the finite-dimensional types over A are precisely, up to interdefinability, the generic types of *meromorphic D-varieties* over A. We leave it to the reader to both articulate precisely, and verify, this claim.

References

[Bays et al. 2017] M. Bays, M. Hils, and R. Moosa, "Model theory of compact complex manifolds with an automorphism", *Trans. Amer. Math. Soc.* **369**:6 (2017), 4485–4516. MR Zbl

[Chatzidakis and Pillay 1998] Z. Chatzidakis and A. Pillay, "Generic structures and simple theories", *Ann. Pure Appl. Logic* **95**:1-3 (1998), 71–92. MR Zbl

[Fischer 1976] G. Fischer, *Complex analytic geometry*, Lect. Notes Math. **538**, Springer, 1976. MR Zbl

[Freitag et al. 2022] J. Freitag, R. Jaoui, and R. Moosa, "When any three solutions are independent", *Invent. Math.* **230**:3 (2022), 1249–1265. MR Zbl

[Hrushovski 1998] E. Hrushovski, "Geometric model theory", pp. 281–302 in *Proceedings of the International Congress of Mathematicians* (Berlin, 1998), vol. I, edited by G. Fischer and U. Rehmann, Deutsche Mathematiker Vereinigung, Berlin, 1998. MR Zbl

[Jaoui and Moosa 2022] R. Jaoui and R. Moosa, "Abelian reduction in differential-algebraic and bimeromorphic geometry", preprint, 2022. To appear in *Ann. Inst. Fourier*. arXiv 2207.07515

[Johnson 2020] W. Johnson, "On the proof of elimination of imaginaries in algebraically closed valued fields", *Notre Dame J. Form. Log.* **61**:3 (2020), 363–381. MR Zbl

[Kim and Pillay 1997] B. Kim and A. Pillay, "Simple theories", *Ann. Pure Appl. Logic* **88**:2-3 (1997), 149–164. MR Zbl

[Kobayashi 1972] S. Kobayashi, *Transformation groups in differential geometry*, Ergebnisse der Math. **70**, Springer, 1972. MR Zbl

[Moosa 2004] R. Moosa, "A nonstandard Riemann existence theorem", *Trans. Amer. Math. Soc.* **356**:5 (2004), 1781–1797. MR Zbl

[Moosa 2005a] R. Moosa, "The model theory of compact complex spaces", pp. 317–349 in *Logic Colloquium* '01 (Vienna, 2001), edited by M. Baaz et al., Lect. Notes Log. **20**, Assoc. Symbol. Logic, Urbana, IL, 2005. MR Zbl

[Moosa 2005b] R. Moosa, "On saturation and the model theory of compact Kähler manifolds", *J. Reine Angew. Math.* **586** (2005), 1–20. MR Zbl

[Moosa 2022] R. Moosa, "Six lectures on model theory and differential-algebraic geometry", preprint, 2022. arXiv 2210.16684

[Moosa and Pillay 2008] R. Moosa and A. Pillay, "Model theory and Kähler geometry", pp. 167–195 in *Model theory with applications to algebra and analysis*, vol. 1, edited by Z. Chatzidakis et al., London Math. Soc. Lect. Note Ser. **349**, Cambridge Univ. Press, 2008. MR Zbl

[Pillay 2000] A. Pillay, "Some model theory of compact complex spaces", pp. 323–338 in *Hilbert's tenth problem*: *relations with arithmetic and algebraic geometry* (Ghent, 1999), edited by J. Denef et al., Contemp. Math. **270**, Amer. Math. Soc., Providence, RI, 2000. MR Zbl

[Pillay and Ziegler 2003] A. Pillay and M. Ziegler, "Jet spaces of varieties over differential and difference fields", *Selecta Math.* (*N.S.*) **9**:4 (2003), 579–599. MR Zbl

[Rebelo 2004] J. C. Rebelo, "Meromorphic vector fields and elliptic fibrations", *Michigan Math. J.* **52**:1 (2004), 33–59. MR Zbl

[Zilber 1993] B. Zilber, "Model theory and algebraic geometry", pp. 202–222 in *Proceedings of the Tenth Easter Conference on Model Theory* (Wendisch Rietz, Germany, 1993), edited by M. Weese and H. Wolter, Seminarberichte **93-1**, Humboldt Universität, Fachbereich Mathematik, Berlin, 1993. MR Zbl

Received 7 Mar 2023. Revised 13 Jul 2023.

RAHIM MOOSA:

rmoosa@uwaterloo.ca
Department of Pure Mathematics, University of Waterloo, Waterloo, ON, Canada

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.285

Revisiting virtual difference ideals

Zoé Chatzidakis and Ehud Hrushovski

In difference algebra, basic definable sets correspond to prime ideals that are invariant under a structural endomorphism. The main idea of an article with Peterzil (*Proc. London Math. Soc.* **85**:2 (2002), 257–311) was that periodic prime ideals enjoy better geometric properties than invariant ideals, and to understand a definable set, it is helpful to enlarge it by relaxing invariance to periodicity, obtaining better geometric properties at the limit. The limit in question was an intriguing but somewhat ephemeral setting called virtual ideals. However, a serious technical error was discovered by Tom Scanlon's UCB seminar. In this text, we correct the problem via two different routes. We replace the faulty lemma by a weaker one that still allows recovering all results of the aforementioned paper for all virtual ideals. In addition, we introduce a family of difference equations ("cumulative" equations) that we expect to be useful more generally. Previous work implies that cumulative equations suffice to coordinatize all difference equations. For cumulative equations, we show that virtual ideals reduce to globally periodic ideals, thus providing a proof of Zilber's trichotomy for difference equations using periodic ideals alone.

Introduction

Boris Zilber developed a geometric description of $\aleph_1$-categorical theories, having a trichotomy at its heart. It is based on the dimension theory of Morley (shown to take finite values by Baldwin), but gives information of a radically new kind than an abstract dimension theory. Intuitively, a model of the theory is coordinatized by geometries that have either a graph-theoretic nature, or derive from linear algebra, or belong to algebraic geometry. Though it is only the minimal definable sets that are described in this way, Zilber (and later others) demonstrated an overwhelming effect on the structure globally.

Zilber conjectured that there is no fourth option. This turned out to be incorrect at the precise level of generality of $\aleph_1$-categorical structures. But it was established with additional hypotheses of a topological nature [11], and moreover

Chatzidakis was partially supported by ValCoMo (ANR-13-BS01-0006). The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (FP7/2007- 2013)/ERC Grant Agreement No. 291111.
MSC2020: 03C60, 12H10.
Keywords: difference ideal, Zilber trichotomy, virtual ideal, model theory.

proved to be meaningful and indeed to capture the nature of structures far beyond strong minimality. Appropriate versions hold for compact complex manifolds, for differentially closed and separably closed fields, for strongly minimal sets interpretable in algebraically closed fields of characteristic zero [1]; the latter closes in characteristic zero a line opened more than thirty years ago by Eugenia Rabinovich, in her Kemerovo PhD with Zilber. The trichotomy is also meaningful for unstable theories: see [12] for the o-minimal case. Many applications depend on the trichotomy, including Zilber's gem [15]. For difference equations, applications to diophantine geometry include [3; 4; 10; 13].

Thanks to Zilber's philosophy, when we made our first steps in the structure of difference equations in [2], we knew in advance what it is that we should aim to prove. The methods were informed by finite-rank stability and the nascent generalization to simplicity. But they also relied strongly on ramification divisors, and thus applied only in characteristic zero. Our approach in [7] to the positive characteristic case thus had to be different.

The trichotomy results of [11] are valid for stable structures with a finite dimension assigned to definable sets, satisfying a "dimension theorem" controlling dimensions of intersections. Now the model companion ACFA of the theory of difference fields is not stable, nor does the geometry of finite-dimensional sets satisfy the dimension theorem: the intersection of two such sets may have unexpectedly low dimension. For instance, the naive intersection of two surfaces in 3-space over the fixed field of the automorphism σ could be two lines interchanged by σ; within the fixed field their intersection point would be the only solution. Both of these pathologies are ameliorated as one relaxes σ to σ^m (going from the equation $\sigma(x) = F(x)$ to $\sigma^m(x) = F^{(m)}(x)$). At the limit, one has a *virtual structure*, defined and studied in [7]; under appropriate conditions, this structure is stable and the dimension theorem is valid. Proving this uses basic ideas from topological dynamics to obtain recurrent points that may not be periodic; see Lemma 2.8 for example. Using a generalization of the Zariski geometries of [11], one can then deduce the trichotomy theorem. The concrete form it takes here allows analyzing any difference equation via a tower of equations over fixed fields and equations of locally modular type.

In 2015, however, Tom Scanlon's Berkeley seminar recognized a problem with a key technical lemma, Lemma 3.7. We show below how to prove a somewhat weaker version of this lemma: where the wrong Lemma 3.7 asserted a unique component through a point, the corrected version, Lemma 2.16, implies that the number of such components is finite, indeed at most the degree of the normalization of the relevant variety in the base. All the main results of the paper remain valid with the same set of ideas, but considerable reorganization is required. One role of the present paper is to provide a lengthy erratum, explaining in detail how this may be done. Parts of this paper are thus technical and need to be read in conjunction with [7]. However, Section 2, which contains the main correction and in particular the key dimension theorem, is self-contained in the sense of quoting some results from [7] but not requiring entering into their proofs.

At the same time, we take the opportunity to present a setting ("cumulative equations") in which the limit structure is equivalent to an ordinary structure, in the sense that the associated algebraic object is an ordinary ring with its periodic ideals, rather than an abstract limit of such rings as in the virtual case. Results of [5] imply that this setting, while not fully general, suffices to coordinatize all difference equations. It may be of interest for other applications, in particular the study of limit structures for more equations that are not necessarily algebraic over equations of SU-rank one.

We expect that a trichotomy theorem can be proved for Zariski geometries based on Robinson structures. This has so far been worked out only in special cases; the most general treatment is contained in the unpublished PhD thesis of Elsner [9]. Consequently, the trichotomy follows from the basic cumulative case alone, though this is not the case for some of the other results: for finer statements such as a description of the fields definable in the limit structures, both in [7] and here, we use additional features of the specific structure.

Let S be a difference ring, generated by a finitely generated ring R. The main idea of [7] was that as n becomes more and more divisible, more σ^n-ideals appear, and their structures become progressively smoother. However there is also a countercurrent at work: the difference subring R_{σ^n} of (S, σ^n) generated by R may become smaller. This double movement leads to technical complexity. If, however, $\sigma(R)$ is contained in the ring generated by R and $\sigma^n(R)$ for any n, this problem does not arise. It is this behavior (slightly generalized to fraction fields) that we refer to as *cumulative*. It turns out that cumulative difference equations still represent all isogeny classes, and allow for considerable simplification.

We are very grateful to Tom Scanlon, his Berkeley group, and especially Alex Kruckman for identifying the error; and to the anonymous referee for the careful reading and suggestions that have considerably improved the text.

Plan of the paper. In Section 1 we mainly recall definitions and notation from [7]. Section 2 contains the proof of Proposition 2.6 of [7], as well as some useful auxiliary results and remarks. The cumulative case is done in the first half, the general case in the second half. Sections 3 and 4 are devoted to rereading [7] and making the necessary changes and adaptations: Section 3 deals with Sections 2 to 4 of [7], and Section 4 with the remainder of the paper.

1. Setting, notation, basic definitions

1.1. ***Setting and notation.*** In what follows, K is a sufficiently saturated existentially closed difference field, containing an algebraically closed difference subfield k_0, and

Ω a $|K|^+$-saturated existentially closed difference field containing K. We always work inside Ω.

If L is a field, then L^s and L^{alg} denote the separable and algebraic closure of the field L.

Conventions. Unless otherwise stated, all difference fields and rings will be *inversive*, i.e., the endomorphism σ is an automorphism; in other words, we take a difference ring to be a commutative ring with a $\mathbb{Z}$-action. Similarly, all difference ideals will be *reflexive*, i.e., if (R, σ) is a difference ring, a σ-ideal of R is an ideal I such that $\sigma(I) = \sigma^{-1}(I) = I$.

If k is a difference field, $X = (X_1, \dots, X_n)$, then $k[X]_\sigma$ denotes the inversive difference domain $k[\sigma^i(X_j) \mid i \in \mathbb{Z}, 1 \leq j \leq n]$ and $k(X)_\sigma$ its field of fractions. Similarly, if a is a tuple in Ω, $k[a]_\sigma$ and $k(a)_\sigma$ denote the inversive difference subring and subfield of Ω generated by a over k. Similar notation is used for difference rings. If a is an n-tuple, then $I_\sigma(a/k) = \{f \in k[X_1, \dots, X_n]_\sigma \mid f(a) = 0\}$. If $k(a)_\sigma$ has finite transcendence degree over k, the *limit degree* of a over k, denoted $\text{ld}(a/k)$ or $\text{ld}_\sigma(a/k)$, is $\lim_{n\to\infty}[k(a, \dots, \sigma^{n+1}(a)) : k(a, \dots, \sigma^n(a))]$.

If A is a subset of a difference ring S, then $(A)_{\sigma^m}$ denotes the (reflexive) σ^m-ideal of S generated by A. If $A \subset \Omega$, then $\text{cl}_\sigma(A)$ denotes the perfect closure of the difference subfield of Ω generated by A, $\text{acl}_\sigma(A)$ the (field-theoretic) algebraic closure of $\text{cl}_\sigma(A)$, and $\text{dcl}_\sigma(A)$ the model-theoretic definable closure of A. If A is a subring of a difference ring S, then A_σ denotes the (inversive) difference subring of S generated by A.

Recall that $\text{acl}_\sigma(A)$ coincides with the model-theoretic algebraic closure $\text{acl}(A)$, and that independence (in the sense of the difference field Ω) of A and B over a subset C coincides with the independence (in the sense of ACF) of $\text{acl}(A)$ and $\text{acl}(B)$ over $\text{acl}(C)$.

If $m \geq 1$, then $\Omega[m]$ denotes the σ^m-difference field (Ω, σ^m). The languages $\mathcal{L}$ and $\mathcal{L}[m]$ are the languages $\{+, -, \cdot, 0, 1, \sigma\}$ and $\{+, -, \cdot, 0, 1, \sigma^m\}$. We view $\mathcal{L}[m]$ as a sublanguage of $\mathcal{L}$, and $\Omega[m]$ as a reduct of Ω. Recall that $\Omega[m]$ is also an existentially closed saturated difference field, by Corollary 1.12 of [2]. If a is a tuple of Ω and k a difference subfield of Ω, then $\text{qftp}(a/k)$ denotes the quantifier-free type of a over k in the difference field Ω, and if $m \geq 1$, then $\text{qftp}(a/k)[m]$ denotes the quantifier-free type of a over k in the difference field $\Omega[m]$. Similarly, if q is a quantifier-free type over k, then $q[m]$ denotes the set of $\mathcal{L}(k)[m]$ quantifier-free formulas implied by q.

Basic and semibasic types.

Definitions 1.2. We consider quantifier-free types $p, q, \dots$ over the algebraically closed difference field k_0, and integers $m, n \geq 1$.

(1) q satisfies (ALGm) if whenever a realizes q, then $\sigma^m(a) \in k_0(a)^{\mathrm{alg}}$.

(2) The *eventual SU-rank* of q, $\mathrm{evSU}(q)$, is $\lim_{m\to\infty} \mathrm{SU}(q[m!])$, where $\mathrm{SU}(q[m!])$ (the SU-rank of $q[m!]$) is computed in the $\sigma^{m!}$-difference field $\Omega[m!]$. For more details, see Section 1 in [7], starting with 1.10. Write $\mathrm{SU}(a/k_0)[n] := \mathrm{SU}(q[n])$, computed in the σ^n-difference field $\Omega[n]$ ($n \geq 1$, a realizing q). If D is a countable union of k-definable subsets of some cartesian power of Ω, then $\mathrm{evSU}(D) = \sup\{\mathrm{evSU}(a/k) \mid a \in D\}$.

(3) $p \sim q$ if and only if for some $m \geq 1$, $p[m] = q[m]$. The $\sim$-equivalence class of p is denoted by $[p]$ and is called a *virtual type*.

(4) $\mathcal{X}_p(K)$ denotes the set of tuples in K which realize $p[m]$ for some $m \geq 1$, and similarly for $\mathcal{X}_p(\Omega)$. We denote by X_p the underlying affine variety, i.e., the Zariski closure of $\mathcal{X}_p(\Omega)$ in affine space.

(5) A *basic* type is a quantifier-free type p over k_0, with evSU-rank 1, which satisfies (ALGm) for some m. Note that if p is basic, so is $p[n]$ for every n.

(6) A *semibasic type* is a quantifier-free type q such that if a realizes q, then there are tuples $a_1, \ldots, a_n \in k_0(a)^{\mathrm{alg}}$ which realize basic types over k_0, are algebraically independent over k_0, and are such that $a \in k_0(a_1, \ldots, a_n)^{\mathrm{alg}}$.

(7) The quantifier-free type q is *cumulative* if for some (any) realization a of q and every $m \geq 1$, $\sigma(a) \in k_0(a, \sigma^m(a))$. Note that this implies that $k_0(a)_\sigma = k_0(a)_{\sigma^m}$ for any $m \geq 1$, and that (ALGm) is equivalent to (ALG1).

Remarks 1.3. Let k be an inversive difference field.

(1) We will often use the following equivalences, for a tuple a:

  (i) $[k(a, \sigma(a)) : k(a)] = \mathrm{ld}(a/k)$.

  (ii) The fields $k(\sigma(a) \mid i \leq 0)$ and $k(\sigma^i(a) \mid i \geq 0)$ are linearly disjoint over $k(a)$.

  (iii) $I_\sigma(a/k)$ is the unique prime σ-ideal of $k[X]_\sigma$ extending the prime ideal $\{f(X, \sigma(X)) \in k[X, \sigma(X)] \mid f(a, \sigma(a)) = 0\}$ of $k[X, \sigma(X)]$ ($|X| = |a|$).

  Note that these equivalent conditions on the tuple a in the difference field Ω also imply the analogous conditions for the tuple a in the difference field $\Omega[m]$ for $m \geq 1$ (use (ii)).

(2) Let P be a prime ideal of $k[X, \sigma(X)]$ (X a tuple of variables) and assume that $\sigma(P \cap k[X]) = P \cap k[\sigma(X)]$. Then P extends to a prime σ-ideal of $k[X]_\sigma$. We will usually use it with the prime ideal $\sigma^{-1}(P)$ of $k[\sigma^{-1}(X), X)]$.

Proof. All these are straightforward remarks; see also Section 1.3 of [5] for the equivalence of (i) and (ii), and Sections 5.6 and 5.2 of [8] for the remaining items. □

1.4. ***Coordinate rings associated to quantifier-free types.*** (See also (3.5) and (3.6) in [7]). Let q be a quantifier-free type over k_0, in the tuple x of variables, fix a realization a of q. The *pair* $(R_q, R_{q,\sigma})$ *of coordinate rings associated to* q is defined as follows: Let $k_0(x)_\sigma$ be the fraction field of $k_0[X]_\sigma/I_\sigma(a/k_0)$, $k_0(x)$ its subfield generated by x over k_0. Then we define the ring $R_q := k_0(x) \otimes_{k_0} K$ and the σ^m-difference ring $R_{q,\sigma^m} := k_0(x)_{\sigma^m} \otimes_{k_0} K$ for $m \geq 1$. We often denote R_q and R_{q,σ^m} by $K\{x\}$ and $K\{x\}_{\sigma^m}$, and define in an analogous way the coordinate rings $k_1\{x\}$ and $k_1\{x\}_{\sigma^m}$ if k_1 is a difference field containing k_0.

Given semibasic types $q_1(x_1), \dots, q_n(x_n)$, we take the tensor product over K of their coordinate rings, and call them the coordinate rings associated to $(q_1, \dots, q_n)$. So, we have

$$R_{(q_1,\dots,q_n)} = K\{x_1\} \otimes_K \cdots \otimes_K K\{x_n\},$$
$$R_{(q_1,\dots,q_n),\sigma^m} = K\{x_1\}_{\sigma^m} \otimes_K \cdots \otimes_K K\{x_n\}_{\sigma^m}.$$

To a semibasic type q, we associate three new pairs of coordinate rings as follows. Say q is realized by a tuple a, and $a_1, \dots, a_n$ are as in the definition of semibasic given above. We let $p_i = \mathrm{qftp}(a_i/k_0)$, $r = \mathrm{qftp}(a_1, \dots, a_n/k_0)$, and $s = \mathrm{qftp}(a, a_1, \dots, a_n/k_0)$. Then we define

$$R_q^1 = R_{p_1} \otimes_K \cdots \otimes_K R_{p_n}, \qquad R_q^2 = R_r, \qquad R_q^3 = R_s,$$
$$R^1_{q,\sigma^m} = R_{p_1,\sigma^m} \otimes_K \cdots \otimes_K R_{p_n,\sigma^m}, \qquad R^2_{q,\sigma^m} = R_{r,\sigma^m}, \qquad R^3_{q,\sigma^m} = R_{s,\sigma^m}.$$

These rings depend on the choice of the tuples $a_1, \dots, a_n$, but we may fix once and for all these tuples. Note that then $R_q^1 \subseteq R_q^2 \subseteq R_q^3 \supseteq R_q$, that R_q^2 is a localization of R_q^1, and that R_q^3 is integral algebraic over R_q^2 and over R_q. Similar statements hold for the associated difference rings. If q is basic, we define $R_q^i = R_q$ and $R^i_{q,\sigma^m} = R_{q,\sigma^m}$. We extend the notation to the more general coordinate rings $R_{(q_1,\dots,q_n)}$.

We say that a coordinate ring R_σ satisfies (ALGm) or is cumulative, if the semibasic types involved in the definition of R_σ all satisfy (ALGm) or are cumulative.

1.5. ***Convention.*** From now on, all quantifier-free types will satisfy (ALGm) for some $m \geq 1$, so that all coordinate rings will satisfy (ALGm).

Definitions 1.6. Let (R, R_σ) be a pair of coordinate rings, as defined above, and S a ring.

(1) Let P be a prime ideal of a ring S. The *dimension* of P, denoted by $\dim(P)$, is the Krull dimension of the ring S/P. If I is an ideal of S, the *dimension* of I, $\dim(I)$, is $\sup\{\dim(P) \mid P \supseteq I,\ P \in \mathrm{Spec}(S)\}$. If $S = R$, then $\dim(P)$ coincides with $\mathrm{tr.deg}_K \mathrm{Frac}(R/P)$.

(2) Let P be a prime ideal of a coordinate ring R_σ. The *virtual dimension of* P, denoted $\mathrm{vdim}(P)$, is $\dim(P \cap R)$. If R_σ satisfies (ALGm), it coincides with $\dim(P \cap R_{\sigma^m})$. Similarly, if I is an ideal of R_σ, then $\mathrm{vdim}(I) = \dim(I \cap R)$.

(3) A *virtual* [*perfect*], [*prime*] *ideal* of R_σ is a [perfect[1]], [prime] (reflexive) σ^m-ideal of R_{σ^m} for some $m \geq 1$.

(4) A [*perfect*], [*prime*] *periodic ideal* of R_σ is a [perfect], [prime] σ^m-ideal I of R_σ for some $m \geq 1$. A priori, not all virtual ideals extend to periodic ideals.

(5) Let I be an ideal of R. We say that I is *pure of dimension d* if all minimal primes over I have dimension d. Let I be an ideal of R_σ. We say that I is *virtually pure of dimension d* if $I \cap R$ is pure of dimension d.

(6) Let I be a virtual ideal of $R_\sigma = K\{x\}_\sigma$. Then $V(I)$ is the subset of $K^{|x|}$ defined by $a \in V(I)$ if and only if for some $m \geq 1$, for each $h \in I \cap R_{\sigma^m}$, viewed as a σ^m-polynomial, we have $h(a, \sigma^m(a), \ldots) = 0$. Thus $V(I)$ stands in bijection with $\bigcup_m \mathrm{Hom}_{\sigma^m}(R_{\sigma^m}/I, K)$, where Hom_{σ^m} refers to ring homomorphisms commuting with σ^m.

Note that if $R_\sigma = R_q$ for some quantifier-free type q, then $V(0)$ is precisely $\mathcal{X}_q(K)$. We call vdim(0) (i.e., the Krull dimension of R) the (*virtual*) *dimension of q*.

2. Existence theorems for periodic ideals

The aim of this section is to give proofs of the results of [7] needed towards the proof of the trichotomy in positive characteristic, and in particular the very important Proposition 2.6 of [7]. We try to follow the plan of [7], and will occasionally refer to it. While the results of Chapter 2 are indeed correct, the problem is that our coordinate rings do not satisfy the required hypotheses. The mistake appears in Lemma 3.7.

Assumptions. The coordinate rings we consider are those associated to tensor products of coordinate rings of semibasic types whose corresponding basic types have virtual dimension e, *for some fixed integer* $e \geq 1$. A typical pair of coordinate rings is denoted (R, R_σ), without reference to the types involved in the construction.

As for types, we declare two virtual prime ideals P, Q equivalent, and write $P \sim Q$, if for some $m \geq 1$, $P \cap R_{\sigma^m} = Q \cap R_{\sigma^m}$. We retain, however, Definition 1.2(3) of virtual prime ideals; the equivalence classes are called virtual prime ideal classes.

Proposition 2.1 (addendum to Proposition 2.4 of [7]). *Let* (R, R_σ) *be a pair of coordinate rings.*

(1) *Let P and Q be virtual prime ideals. If $V(P) = V(Q)$, then $P \sim Q$.*

(2) *Let P be a prime σ^m-ideal of R_{σ^m}. Then for some $\ell > 0$, P extends to a prime σ^ℓ-ideal Q of R_σ. In particular, since $V(Q) = V(P)$, this shows that every set defined by a virtual prime ideal is also defined by a periodic prime ideal*

[1] A σ-ideal I of a difference ring R is perfect if whenever $a^n\sigma(a) \in I$, then $a \in I$.

of R_σ; i.e., every prime periodic ideal of R_{σ^m} is equivalent to a prime periodic ideal of R_σ.

Proof. (1) We may assume that P and Q are prime σ-ideals and that R satisfies (ALG1). Choose a (small) subfield k_1 of K such that for any $m \geq 1$, $P \cap R_{\sigma^m}$ and $Q \cap R_{\sigma^m}$ are generated by their intersection with $k_1\{x\}_{\sigma^m}$ (x the variables of R). By saturation of K, it contains a point a which is a generic point of $V(P)$ over k_1, i.e., with $\text{tr.deg}(k_1(a)/k_1) = \dim(P)$. Then a is in $V(Q)$, whence $\dim(Q) \geq \dim(P)$, and the symmetric argument tells us that these dimensions are equal, and that a is a generic of $V(Q)$ over k_1. Let ℓ be divisible by m and such that $P \cap R_{\sigma^\ell}$ and $Q \cap R_{\sigma^\ell}$ are prime σ^ℓ-ideals contained in $(x-a)_{\sigma^\ell}$. Then $I_{\sigma^\ell}(a/k_1) = P \cap k_1\{x\}_{\sigma^\ell} = Q \cap k_1\{x\}_{\sigma^\ell}$, which shows that $P \sim Q$.

(2) Let $\varphi : R_{\sigma^m} \to \Omega$ be a K-homomorphism of σ^m-difference rings with kernel P. If $p_1(x_1), \dots, p_n(x_n)$ are the semibasic types associated to R_σ, then

$$R_\sigma = k_0(x_1)_\sigma \otimes_{k_0} \cdots \otimes_{k_0} k_0(x_n)_\sigma \otimes_{k_0} K,$$

and R_{σ^m} corresponds to the subring $k_0(x_1)_{\sigma^m} \otimes_{k_0} \cdots \otimes_{k_0} k_0(x_n)_{\sigma^m} \otimes_{k_0} K$. Our map φ is entirely determined by its restrictions to each of the factors of the tensor product, and for $i = 1, \dots, n$, we let φ_i denote the restriction of φ to $k_0(x_i)_{\sigma^m}$. Since $k_0(x)_\sigma$ is finitely generated over $k_0(x)_{\sigma^m}$, Proposition 1.12(3) of [7] gives that for some $\ell > 0$ divisible by m, the σ^ℓ-embeddings $\varphi_i : k_0(x_i)_{\sigma^m} \to \Omega$ extend to σ^ℓ-embeddings $\psi_i : k_0(x)_\sigma \to \Omega$ for $i = 1, \dots, n$. Then define $\psi = \psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n \otimes \text{id}_K$, and take $Q = \ker \psi$. □

Lemma 2.2. *Let R_σ be a coordinate ring, and $S_\sigma = R[c]_\sigma$ a difference ring, with $S = R[c]$ integral algebraic (and finitely generated) over R. If P is a prime σ-ideal of R_σ, then for some $\ell \geq 1$, $P \cap R_{\sigma^\ell}$ extends to a prime σ^ℓ-ideal of S_{σ^ℓ}.*

Proof. Replacing σ by σ^m for some m, we may assume that R_σ satisfies (ALG1).

Claim. *There is $m \geq 1$ such that for any $\ell \geq 1$, if $R' = R[\sigma(R), \dots, \sigma^m(R)]$, then $P \cap R'_{\sigma^\ell}$ is the unique prime σ^ℓ-ideal of R'_{σ^ℓ} which extends $P \cap R'[\sigma^\ell(R')]$.*

Proof of claim. Indeed, take $a \in \Omega$ such that $\text{Frac}(R_\sigma/P) \simeq K(a)_\sigma$ and m such that $[K(a, \dots, \sigma^{m+1}(a)) : K(a, \dots, \sigma^m(a))] = \text{ld}(a/K)$. Then if $b = (a, \dots, \sigma^m(a))$, we have $\text{ld}(b/K) = \text{ld}(a/K)$ and for $\ell \geq 1$, $\text{ld}_{\sigma^\ell}(b/k_0) = [K(b, \sigma^\ell(b)) : K(b)]$.

The claim now follows by the equivalences given in Remarks 1.3(1). □

For $n \geq 0$, let $R(n)$ and $S(n)$ denote the subrings of R_σ and S_σ generated respectively by $\sigma^i(R)$ and $\sigma^i(S)$, $-n \leq i \leq n$. Then each $S(n)$ is Noetherian, integral algebraic over $R(n)$, $S_\sigma = \bigcup_{n\in\mathbb{N}} S(n)$, and we have a natural map

$$\text{Spec}(S_\sigma) \to \prod_{n\in\mathbb{N}} \text{Spec}(S(n)).$$

For each $n \in \mathbb{N}$, the set X_n of prime ideals of $S(n)$ which extend $P \cap R(n)$ is finite and nonempty, and the natural map $\mathrm{Spec}(S(n+1)) \to \mathrm{Spec}(S(n))$ sends X_{n+1} to X_n. Hence $X := \varprojlim X_n$ is a closed, compact, nonempty subset of $\prod_{n\in\mathbb{N}} X_n$, and is the set of prime ideals of S_σ which extend P. As each X_n is finite, and the set X is stable under the (continuous) action of σ on $\mathrm{Spec}(S_\sigma)$, X contains a recurrent point Q. Let m be given by the claim, and consider $S(m)$. Then for some $\ell \geq 1$, we have $\sigma^\ell(Q) \cap S(m) = Q \cap S(m)$, and therefore, using Remarks 1.3(2), there is a prime σ^ℓ-ideal Q' of $S(m)_{\sigma^\ell}$ such that

$$Q' \cap S(m)[\sigma^{-\ell}(S(m))] = Q \cap S(m)[\sigma^{-\ell}(S(m))].$$

As Q contains $P \cap R'[\sigma^{-\ell}(R')]$ and has the same dimension, by the claim Q' must extend $P \cap R'_{\sigma^\ell}$, and therefore also $P \cap R_{\sigma^\ell}$. □

Remark 2.3. A consequence of our hypothesis on the dimension of the basic types is as follows: Let P be a virtual prime ideal of R_σ. Then $\dim(P \cap R)$ is divisible by e. Indeed, choose m such that $P \cap R_{\sigma^m}$ is a prime σ^m-ideal of R_{σ^m} and R_σ satisfies (ALGm). We may assume that $m = 1$. We use the notation and definition of Section 1.4, and recall that R^3 is finite integral algebraic over R. Thus, by Lemma 2.2, $P \cap R_\sigma$ extends to a periodic prime ideal of R^3_σ. This means that $\mathrm{Frac}(R_\sigma/P \cap R_\sigma)$ is equi-algebraic over K to a difference field which is generated over K by realizations of basic types of dimension e. Since basic types have evSU-rank 1, these realizations may be taken independent, and therefore $\mathrm{tr.deg}_K(\mathrm{Frac}(R_\sigma/P \cap R_\sigma))$ is a multiple of e, so that $\dim(P \cap R_\sigma)$ is a multiple of e. As R_σ is integral algebraic over R, $\dim(P \cap R)$ is a multiple of e.

The basic cumulative case. We now prove some results in the particular case when our coordinate rings are tensor products of coordinate rings of *basic cumulative* types; this assumption holds until Proposition 2.10. The proof in the general case follows the same lines, but is slightly more involved.

Note that the assumptions imply that all coordinate rings satisfy (ALG1), that all virtual ideals are periodic, and that $\sim$ coincides with equality.

Lemma 2.4. *Let I be an ideal of R of dimension d. Then there are only finitely many periodic prime ideals of R_σ which contain I and are of dimension d.*

Proof. A prime ideal of R_σ which contains I and is of dimension d must extend a prime ideal P of R of dimension d containing I. As R is Noetherian, there are only finitely many such prime ideals, and we may therefore assume that $I = P$ is prime, and extends to a periodic prime ideal of R_σ.

Then Proposition 3.10 of [7], together with Proposition 2.1, gives the result. □

Corollary 2.5. *Let I be an ideal of R_σ of dimension d. Then there are only finitely many periodic prime ideals of R_σ which contain I and are of dimension d.*

Proof. Such an ideal contains in particular $I \cap R$. The result then follows from Lemma 2.4. □

Corollary 2.6. *Let I be an ideal of R_σ of dimension d. Then there are periodic prime ideals $P_1, \ldots, P_s$ of R_σ of dimension d, and a finite subset F of I, such that if P is a periodic prime ideal of R_σ which contains F and is of dimension d, then $V(P) = V(P_i)$ for some i.*

Proof. By Lemma 2.4, if F is a finite subset of R_σ which generates an ideal of dimension d and $\mathrm{per}(F)$ denotes the set of prime periodic ideals of R_σ containing F and of dimension d, then $\mathrm{per}(F)$ is finite. Take a sufficiently large finite F such that $\mathrm{per}(F) = \mathrm{per}(I)$. □

Lemma 2.7. *Let I be a periodic ideal of R_σ of dimension d. Then I is contained in a periodic prime ideal of R_σ of dimension d.*

Proof. We may assume that $I = \sigma(I)$. Let $F \subset I$ and $P_1, \ldots, P_s$ be given by Corollary 2.6. Let X be the set of prime ideals of R_σ of dimension d containing I, and for $n \in \mathbb{N}$, let $R(n)$ be the subring of R_σ generated by $\sigma^i(R)$, $-n \le i \le n$, and X_n be the set of prime ideals of $R(n)$ containing $I \cap R(n)$ and of dimension d. Each X_n is finite and nonempty, and we have natural maps $X_{n+1} \to X_n$. Hence, $X = \varprojlim X_n$ is nonempty and compact. The automorphism σ acts continuously on X, and therefore has a recurrent point Q. Let n be such that $R(n)$ contains F. Then for some $m > 0$, we have $Q \cap R(n) = \sigma^m(Q) \cap R(n)$. By Remarks 1.3(2), there is a prime σ^m-ideal Q' of $R(n)_{\sigma^m}$ which extends $Q \cap R(n)[\sigma^{-m}(R(n))]$. But $R(n)_{\sigma^m} = R_\sigma$, and because Q' contains F and has dimension d, it must contain I. □

Lemma 2.8. *Let I be a periodic ideal of R_σ, with $I \cap R$ pure of dimension d. Then there are periodic prime ideals $P_1, \ldots, P_s$ of virtual dimension d such that $V(I) = V(P_1) \cup \cdots \cup V(P_s)$.*

Proof. We already know by Lemma 2.4 (and Proposition 2.1) that $V(I)$ has only finitely many irreducible components of dimension d, say $V(P_1), \ldots, V(P_s)$. It therefore suffices to show that every point of $V(I)$ is in one of these components.

Assume this is not the case. Let $a \in V(I)$, and $m \ge 1$ such that I is a σ^m-ideal and $Q = (x - a)_{\sigma^m} \supseteq I$. Without loss of generality, $m = 1$. For $n \in \mathbb{N}$, let $R(n)$ be the subring of R_σ generated by the rings $\sigma^i(R)$, $-n \le i \le n$. Then for each $n \in \mathbb{N}$, the ideal $I \cap R(n)$ is pure of dimension d, and therefore, the set X_n of prime ideals P of $R(n)$ of dimension d containing $I \cap R(n)$ and contained in Q is finite and nonempty. Moreover, if $P \in X_{n+1}$, then $P \cap R(n) \in X_n$. Hence, the compact subset $X = \varprojlim X_n$ of $\mathrm{Spec}(R_\sigma)$ is nonempty. It is the set of prime ideals of R_σ of dimension d, containing I and contained in Q. Let F be given by Corollary 2.6, and n such that $F \subset R(n)$ and Q does not contain any of the $P_i \cap R(n)$. As σ acts continuously on the compact set X, X has a recurrent point, say P. Then for some $m \geq 1$, $P \cap R(n) = \sigma^m(P) \cap R(n)$. As in the proof of Lemma 2.7, there is a prime σ^m-ideal P' of R_σ which extends $P \cap R(n)[\sigma^{-m}(R(n))]$, and therefore has dimension d, contains I and is not in the finite set $\{P_1, \ldots, P_s\}$. This gives us the desired contradiction. □

We define a topology on V, taking the closed sets to be the sets $V(I)$. (It is easy to see that the sets $V(I)$ are closed under intersections and under finite unions.) Then when s is taken minimal in Lemma 2.8, the $V(P_i)$ are the *irreducible components* of $V(I)$.

Lemma 2.9. *Write $R_\sigma = K\{x_1\} \otimes_K \cdots \otimes_K K\{x_m\}$, with $m \geq 2$. Let P be a prime σ-ideal of R_σ, and let Q be the ideal $Q = (x_1 - x_2)_\sigma$ corresponding to the diagonal on* Spec $K\{x_1\} \times$ Spec $K\{x_2\}$, *i.e., generated by the $x_{1,j} - x_{2,j}$. Then either $Q \subseteq P$, or every irreducible component of $V(P) \cap V(Q)$ has dimension* $\dim(P) - e$.

Proof. Assume $Q \nsubseteq P$, and consider the σ-ideal $I = P + Q$. Note that since Q is generated by elements of R, at least one of them is not in P. Thus $I \cap R$ is strictly bigger than $P \cap R$, so each component of $I \cap R$ has dimension $< \dim(P)$.

Let $R(n)$ be the subring of R_σ generated by $\sigma^i(R)$, $-n \leq i \leq n$, for $n \in \mathbb{N}$. Then each $R(n)$ is a localization of the affine coordinate ring of a smooth variety. (In our construction, *all* proper subvarieties defined over k_0, including the singular locus, were localized away. See the discussion in (5.18) of [7].)

Hence the dimension theorem holds: since $Q \cap R(n)$ has codimension e, all minimal prime ideals over $P \cap R(n) + Q \cap R(n)$ have dimension $\geq \dim(P) - e$.

Since R is Noetherian, $I \cap R$ is finitely generated. Any finite set of elements of $I \cap R$ must already belong to $P \cap R(n) + Q \cap R(n)$ for some n. Since $R(n)$ is integral over R, and the components of $P \cap R(n) + Q \cap R(n)$ have dimension $\geq \dim(P) - e$, it follows that every minimal prime of $I \cap R$ has dimension $\geq \dim(P) - e$. (The image of an irreducible variety under a morphism with finite fibers is an irreducible variety of the same dimension.)

In particular, I has dimension $\delta \geq \dim(P) - e$. By Lemma 2.7 some periodic prime ideal P' containing I has dimension δ; by Remark 2.3, δ as well as $\dim(P)$ must be a multiple of e. We saw that $\delta < \dim(P)$, so the only choice is $\delta = \dim(P) - e$. Thus $I \cap R$ is pure of dimension $\dim(P) - e$. Hence Lemma 2.8 applies, and shows that the components $V(P_1), \ldots, V(P_n)$ of $V(I)$ all have dimension exactly $d - e$. □

Proposition 2.10. *Let P and Q be periodic prime ideals of R_σ. Then every irreducible component of $V(P) \cap V(Q)$ has dimension* $\geq \dim(P) + \dim(Q) - \dim(0)$; *it is determined by a periodic prime ideal of R_σ intersecting R in minimal prime ideals over* $(P \cap R) + (Q \cap R)$.

Proof. This can be deduced from Lemma 2.9 by reduction to an intersection with the diagonal Δ (identifying $V(P) \cap V(Q)$ with $P \times Q \cap \Delta$). □

The general case. The results in the cumulative case extend easily to the general case, in most cases simply replacing equality of ideals by the equivalence relation $\sim$. The fact that we consider also coordinate rings of semibasic types makes things a little more complicated, but Lemma 2.2 is of use. Also, Proposition 2.1 allows us to juggle between periodic and virtual ideals. Recall our assumptions:

> (R, R_σ) is a tensor product of coordinate rings of semibasic types, and all associated basic types have virtual dimension e.

Lemma 2.11. *Let I be an ideal of R, of dimension d. Then, up to $\sim$, there are only finitely many virtual prime ideals of R_σ which contain I and are of virtual dimension d.*

Proof. We may assume that R_σ satisfies (ALG1). Then a prime ideal of R_σ which contains I and is of virtual dimension d must extend a prime ideal P of R of dimension d containing I. As R is Noetherian, there are only finitely many such prime ideals, and we may therefore assume that $I = P$ is prime, and extends to a virtual prime ideal of R_σ.

Let us first assume that the semibasic types involved in R_σ are all basic. Then Proposition 3.10 of [7], together with Proposition 2.1, gives us the result.

Let us now do the general case. We consider the rings R^i introduced in Section 1.4. Recall that $R^1 \subseteq R^2 \subseteq R^3 \supseteq R$. As R^3_σ is integral algebraic over R_σ, and satisfies (ALG1), Lemma 2.2 tells us that any virtual prime ideal of R_σ extends to a virtual prime ideal of R^3_σ. On the other hand, there are only finitely many prime ideals of R^3 which extend P, so we may assume that $R = R^3$, $R_\sigma = R^3_\sigma$.

The first case gives us that $P \cap R^1$ extends to finitely many prime virtual ideals of R^1_σ, up to $\sim$, and by Proposition 2.1, we may assume they are periodic. As R^2 and R^2_σ are localizations of R^1 and R^1_σ, respectively, a periodic prime ideal of R^1_σ extends to at most one (periodic) prime ideal of R^2_σ. Say Q is a prime σ^ℓ-ideal of $R^2_{\sigma^\ell}$ which extends $P \cap R^2$. Then there are only finitely many prime ideals of $R^2_{\sigma^\ell}[R^3]$ which extend Q, and by Lemma 3.9 of [7], to each of these corresponds at most one (up to $\sim$) virtual ideal of R^3_σ. Hence, up to $\sim$, there are only finitely many virtual ideals of R^3_σ extending P. □

Corollary 2.12. *Let I be an ideal of R_σ of virtual dimension d. Then, up to $\sim$, there are only finitely many virtual prime ideals of R_σ of virtual dimension d and which contain $I \cap R_{\sigma^m}$ for some $m > 0$.*

Proof. Such an ideal contains in particular $I \cap R$. The result follows from Lemma 2.11. □

Corollary 2.13. *Let I be an ideal of R_σ of virtual dimension d. Then there are periodic prime ideals $P_1, \dots, P_s$ of R_σ of virtual dimension d, and a finite subset*

F of I, such that if P is a periodic prime ideal which contains F and is of virtual dimension d, then $V(P) = V(P_i)$ for some i.

Proof. By Lemma 2.11, if F is a finite subset of R_σ which generates an ideal of dimension d and $\mathrm{per}(F)$ denotes the set of prime periodic ideals of R_σ containing F and of dimension d, then $\mathrm{per}(F)/\sim$ is finite. Take a sufficiently large finite F such that $\mathrm{per}(F)/\sim = \mathrm{per}(I)/\sim$. □

2.14. *Warning.* This set F is not necessarily contained in R, nor in $\bigcap_m R_{\sigma^m}$, unless R_σ is cumulative.

We will need a version of Lemma 2.8 without the purity assumption. We claim a weaker conclusion, namely that $V(I)$ is contained in some $V(P_i)$ of maximal dimension.

Lemma 2.15. *Let I be a virtual ideal of R_σ of virtual dimension d. Then there are $m \geq 1$ and a prime σ^m-ideal of R_{σ^m} of dimension d which contains $I \cap R_{\sigma^m}$.*

Proof. We may assume that $I = \sigma(I)$, and that R_σ satisfies (ALG1). Let $F \subset I$ be given by Corollary 2.13. Let X be the set of prime ideals of R_σ of dimension d containing I, and for $n \in \mathbb{N}$, let $R(n)$ be the subring of R_σ generated by $\sigma^i(R)$, $-n \leq i \leq n$, and X_n be the set of prime ideals of $R(n)$ containing $I \cap R(n)$ and of dimension d. Each X_n is finite, nonempty, and we have natural maps $X \to \prod_{n\in\mathbb{N}} X_n$ and $X_{n+1} \to X_n$. The automorphism σ acts continuously on the compact set X, and therefore has a recurrent point Q. Let n be such that $R(n)$ contains F. Then for some $m > 0$, we have $Q \cap R(n) = \sigma^m(Q) \cap R(n)$. By Remarks 1.3(2), there is a prime σ^m-ideal Q' of $R(n)_{\sigma^m}$ which extends $Q \cap R(n)[\sigma^{-m}(R(n))]$. Applying Proposition 2.1 to $R(n)_{\sigma^m}$, we obtain a prime σ^ℓ-ideal Q'' of R_σ which extends Q'; then Q'' contains F and has dimension d. □

Lemma 2.16 (correct version of Lemma 3.7 in [7]). *Let R be a domain which is integrally closed, k a subfield of R, and k_1 an algebraic extension of k. Let $S = k_1 \otimes_k R$. Let Q be a prime ideal of S.*

(1) *There is a unique prime ideal of S which intersect R in* (0) *and is contained in Q.*

(2) *If P' is a prime ideal of S which intersects R in* (0) *and if k_1 is separably algebraic over k, then S/P' is integrally closed.*

Proof. For both (1) and (2), we may assume that S is finitely generated over R, i.e., that k_1 is a finite extension of k. Furthermore, observe that if $b \in S$, then b^{p^n} belongs to the subring $(k_1 \cap k^s) \otimes_k R$ of S for some n, and that a prime ideal P of S contains b if and only if its intersection with $(k_1 \cap k^s) \otimes_k R$ contains b^{p^n}, i.e., the restriction map $\mathrm{Spec}(S) \to \mathrm{Spec}((k_1 \cap k^s) \otimes_k R)$ is a bijection. We may therefore assume that k_1 is separably algebraic over k of the form $k[a]$ for some $a \in k_1$.

Let $f(T)$ be the minimal monic polynomial of a over k and consider its factorization $\prod_{i=1}^m g_i(T)$ over $\mathrm{Frac}(R)$ into monic irreducible polynomials. Because R is integrally closed, all $g_i(T)$ are in $R[T]$ (see, e.g., Theorem 4, Chapter V, §3 in [14]). Moreover, since f is separable, their coefficients are actually in the subfield $R \cap k^s$ of R, and if $i \neq j$, then $(g_i(T), g_j(T)) = (1)$. Thus any prime ideal of S, and in particular Q, contains one and only one of the elements $g_i(a)$, and the ideal of S generated by $g_i(a)$ is prime. (For this last assertion, use the fact that $g_i(T)$ is irreducible over $\mathrm{Frac}(R)$, and that $S \simeq R[T]/f(T)$). This shows (1).

For (2), viewing R as the coordinate ring of an affine variety V over k, we know that V is normal. A minimal prime ideal of S corresponds therefore to an irreducible component of the (nonirreducible) variety V_{k_1}, and as the property of normality is a local property, each component of V_{k_1} is normal, i.e., with P' as above, S/P' is integrally closed. Here we are using the fact that k_1/k is separable, so that the map $\mathrm{Spec}(k_1) \to \mathrm{Spec}(k)$ is étale and if k_1/k is finite, then S is a product of domains.

The fact that R is not necessarily finitely generated over K is not important: it is a union of finitely generated K-algebras which are integrally closed. □

Proposition 2.17. *Let (R, R_σ) be a pair of coordinate rings associated to semibasic types satisfying* (ALG1). *Then (R, R_σ) satisfies the following: if Q is a prime ideal of R_σ and if P is a prime ideal of R which is contained in $Q \cap R$, then there are only finitely many prime ideals of R_σ which extend P and are contained in Q.*

Proof. Let $Q \subset R_\sigma = S$ be a prime ideal, and let P be a prime ideal of R such that $P \subseteq Q \cap R$. Let us first assume that R/P is integrally closed. Let $(x_1, \dots, x_n)$ be the coordinates corresponding to R, i.e., $R = K\{x_1\} \otimes_K \cdots \otimes_K K\{x_n\}$ and $K\{x_i\} = k_0(x_i) \otimes_{k_0} K$. Then

$$S = \big(\cdots((R \otimes_{K\{x_1\}} K\{x_1\}_\sigma) \otimes_{K\{x_2\}} K\{x_2\}_\sigma) \cdots \otimes_{K\{x_n\}} K\{x_n\}_\sigma\big).$$

We know that each $K\{x_i\}_\sigma$ is integral algebraic over $K\{x_i\}$ (by (ALG1)). However, it may not be separably integral algebraic. So, we consider instead the ring

$$S' = \big(\cdots(R \otimes_{K\{x_1\}} (K\{x_1\}_\sigma \cap K\{x_1\}^s)) \otimes_{K\{x_2\}} \cdots \otimes_{K\{x_n\}} (K\{x_n\}_\sigma \cap K\{x_n\}^s)\big).$$

If $b \in S$, some p^m-th power of b lies in S', so that any prime ideal of S' extends uniquely to a prime ideal of S. It therefore suffices to prove the result for S'.

Applying Lemma 2.16 to $k = K\{x_1\}$ and $S_1 = R \otimes_{K\{x_1\}} (K\{x_1\}_\sigma \cap K\{x_1\}^s)$, we obtain that there is a unique prime ideal P_1 of S_1 which extends P and is contained in $Q \cap S_1$. Furthermore, S_1/P_1 is integrally closed. Iterate the reasoning to obtain that there is a unique prime ideal P_n of S' which extends P and is contained in Q (and furthermore, S'/P_n is integrally closed).

In the general case, let A be the integral closure of R/P. Because R/P is a localization of a finitely generated K-algebra, it follows that A is a finite R/P-module (see Theorem 9, Chapter V, §4 of [14]; observe also that a localization of an integrally closed domain is integrally closed), and is integral algebraic over R/P. So the map $\mathrm{Spec}(A) \to \mathrm{Spec}(R/P)$ is finite, with fibers of size at most g for some g. Hence, the prime ideal Q/PS of S/PS has exactly s extensions $Q_1, \dots, Q_s$ to $\widetilde{S} = (S/PS) \otimes_{R/P} A$, for some s with $1 \le s \le g$. Let P' be a prime ideal of S extending P and contained in Q; then P' contains PS, and therefore P'/PS extends to a prime ideal Q' of $\widetilde{S}$. This Q' must be contained in one of the Q_i. By the first case, this determines Q' uniquely, and therefore also P'. Hence P has at most s extensions to prime ideals of R_σ which are contained in Q. □

Lemma 2.18. *Let I be a virtual perfect ideal of R_σ, with $I \cap R$ pure of dimension d. Then there are periodic prime ideals $P_1, \dots, P_s$ of virtual dimension d such that $V(I) = V(P_1) \cup \dots \cup V(P_s)$.*

Proof. We already know, by Lemma 2.11, that $V(I)$ has only finitely many irreducible components of dimension d. It therefore suffices to show that every point of $V(I)$ is in one of these components. Let $a \in V(I)$, and $m \ge 1$ such that R_σ satisfies (ALGm), $I \cap R_{\sigma^m}$ is a perfect σ^m-ideal and $Q = (x-a)_{\sigma^m} \supseteq I \cap R_{\sigma^m}$. We work in R_{σ^m}, so without loss of generality, $m = 1$. For $n \in \mathbb{N}$, let $R(n)$ be the subring of R_σ generated by the rings $\sigma^i(R)$, $-n \le i \le n$. Then for each $n \in \mathbb{N}$, the ideal $I \cap R(n)$ is pure of dimension d, and therefore, the set X_n of prime ideals P of $R(n)$ of dimension d containing $I \cap R(n)$ and contained in Q is finite and nonempty. Moreover, if $P \in X_{n+1}$, then $P \cap R(n) \in X_n$. Hence, the compact subset $X = \varprojlim X_n$ of $\mathrm{Spec}(R_\sigma)$ is nonempty. It is the set of prime ideals of R_σ of dimension d, containing I and contained in Q. If $P \in X$, then $P \cap R$ belongs to the finite set X_0; hence, by Lemma 2.16, X is finite. On the other hand, X is stable under the (continuous) action of σ, because I and Q are σ-ideals. Hence, for some ℓ, σ^ℓ is the identity on X, i.e., all ideals in X are prime σ^ℓ-ideals. □

Proposition 2.19 (Proposition 2.6 in [7]). *Let $(R, R_\sigma) \in \mathcal{R}$ be a pair of coordinate rings, and let P_1, P_2 be two virtual prime ideals of R_σ. Then $V(P_1) \cap V(P_2) = V(I)$ for some virtual perfect ideal I. The irreducible components of $V(P_1) \cap V(P_2)$ correspond to virtual prime ideals Q_i with $Q_i \cap R$ minimal prime containing $P_1 \cap R + P_2 \cap R$.*

Proof. We may assume that R_σ satisfies (ALG1), and that P_1 and P_2 are prime σ-ideals. (In fact, at every stage of the proof, we allow ourselves to replace R_σ by R_{σ^m} so that our ideals remain σ-ideals, and without explicitly saying so). For the first assertion, it suffices to show that $V(P_1) \cap V(P_2)$ has only finitely many irreducible components: if these are of the form $V(Q_i)$, $i = 1, \dots, s$, for Q_i a prime σ^m-ideal of R_{σ^m}, then one takes $I = \bigcap_{i=1}^s Q_i$, a perfect σ^m-ideal of R_{σ^m} (which contains $P_1 \cap R_{\sigma^m} + P_2 \cap R_{\sigma^m}$).

If $V(P_1) \cap V(P_2) = \varnothing$ then there is nothing to prove, so we assume it is nonempty. The elements of $V(P_1) \cap V(P_2)$ are in correspondence with the elements of $(V(P_1) \times V(P_2)) \cap \Delta$, where the corresponding pair of coordinate rings is $(R_\sigma \otimes_K R_\sigma, R \otimes_K R)$, and Δ denotes the diagonal of the underlying ambient set $V(0) \times V(0)$. The same observation holds at the level of the Zariski closures. We therefore replace P_1 by the ideal P of $R_\sigma \otimes_K R_\sigma$ generated by $P_1 \otimes 1 + 1 \otimes P_2$, and P_2 by the ideal corresponding to Δ, i.e., the ideal $I(\Delta)$ of $R_\sigma \otimes_K R_\sigma$ generated by all $a \otimes 1 - 1 \otimes a$, for $a \in R_\sigma$. Write R_σ as the tensor product over K of the rings $K\{x_i\}_\sigma$, $i = 1, \ldots, n$, with $K\{x_i\}$ associated to the semibasic type q_i. Then $\Delta = \bigcap \Delta_i$, where $\Delta_i \subset V(0) \times V(0)$ is defined by $x_i = x_i'$ inside

$$S_\sigma = (K\{x_1\}_\sigma \otimes_K \cdots \otimes_K K\{x_n\}_\sigma) \otimes_K (K\{x_1'\}_\sigma \otimes_K \cdots \otimes_K K\{x_n'\}_\sigma).$$

It then suffices to show the result for $P + I(\Delta_1)$, then for each $P' + I(\Delta_2)$ where P' is a prime periodic ideal minimal containing $P + I(\Delta_1)$, etc.

Let us first assume that q_i is basic and that P does not contain $I(\Delta_i)$. The proof is very similar to the proof of Lemma 2.9, with small changes. Let $S = R \otimes_K R$, $S_\sigma = R_\sigma \otimes_K R_\sigma$, and $S(n) \subset S_\sigma$ the subring generated by $\sigma^i(S)$, $-n \le i \le n$, for $n \in \mathbb{N}$. Reasoning as in the proof of Lemma 2.9, all minimal prime ideals over $P + I(\Delta_i)$ have dimension $\ge \dim(P) - e$. By Lemma 2.15, $P + I(\Delta_i)$ is contained in a prime periodic ideal P' of dimension $\dim(P + I(\Delta_i))$. By Remark 2.3, $\dim(P + I(\Delta_i))$ must be a multiple of e, and this implies it must equal $\dim(P) - e$. Hence all irreducible components of $V(P + I(\Delta_i))$ have dimension $\dim(P) - e$.

Note that the minimal virtual prime ideals containing $P + I(\Delta_i)$ do indeed extend minimal prime ideals over $P \cap S + I(\Delta_i) \cap S$, since they have the same dimension.

We now do the general case. As R_σ^3 is integral algebraic over R_σ, we may assume that $R_{q_i} = R_{q_i}^3$, $R_{q_i,\sigma} = R_{q_i,\sigma}^3$, by Lemma 2.2. Write the variables of q_i as $(y, y_1, \ldots, y_r)$. Then $I(\Delta_i)$ is the intersection of the r σ-ideals

$$(y_1 - y_1')_\sigma, (y_2 - y_2')_\sigma, \ldots, (y_r - y_r', y - y')_\sigma.$$

The first $r - 1$ of these ideals have dimension $\text{tr.deg}_K(S) - e$ in S_σ; for the last one, work inside $S_\sigma/(y_1 - y_1', y_2 - y_2', \ldots, y_{r-1} - y_{r-1}')_\sigma$. Then the minimal prime σ-ideals over $I(\Delta_i)/(y_1 - y_1', y_2 - y_2', \ldots, y_{r-1} - y_{r-1}')_\sigma$ all have dimension $\text{tr.deg}_K(R_\sigma)$. Apply the first case to these ideals to conclude. □

Corollary 2.20 (the dimension theorem [7, (4.16)]). *Let P_1 and P_2 be virtual prime ideals of R_σ, and let n be the* evSU*-rank of $V(0)$ (i.e., there are exactly n basic types which are associated to R_σ). Then all nonempty irreducible components of $V(P_1) \cap V(P_2)$ have evSU-rank* $\ge (\dim(P_1) + \dim(P_2))/e - n$.

3. Going through Sections 2, 3 and 4 of [7]

We describe which of the results of these three sections remain true without changes, which ones are false or unnecessary, and which ones need to be repaired. Note that while our coordinate rings are not "friendly" (because they do not satisfy ($*1$)), the assumption we make on the semibasic types considered are usually slightly stronger than those made in the paper. Unless preceded by "the present", references are to results in [7].

Section 2. We gave up on the idea of finding a general setting (a modified version of friendliness satisfied by our coordinate rings) in which one would be able to prove the dichotomy theorem, and so in all the results, the hypotheses of friendliness should be replaced by our hypotheses on semibasic types: the associated basic types all have dimension e.

Notation and definitions are given in more details in (2.1) and (2.2), as well as some examples. Proposition 2.4 states the basic results on the duality between sets $V(I)$ and virtual ideals.

Proposition 2.6 is the present Proposition 2.19. The proof of Proposition 2.8 goes through verbatim.

Section 3. Paragraphs (3.1) to (3.6) are definitions and notation.

Lemma 3.7 is *false*. The correct version is given by the present Lemma 2.16(1), but it is not enough to prove ($*1$) for our coordinate rings. Thus Proposition 3.8 is false as well.

However, the proofs of Lemma 3.9 and Proposition 3.10 go through without change (except for a typo on line 4 of the proof of 3.10: it should be the inverse image of $Q \cap K[x_1, \dots, x_r]_\sigma$).

Theorem 3.11 is implied by the present Corollary 2.12.

Proposition 3.12 goes through verbatim (note that the claim is the present Remark 2.3). Note also that once more, Proposition 2.6 (the present Proposition 2.19) is instrumental.

Section 4. Paragraph (4.1) consists of definitions and notation.

Proposition 4.2 remains true, but the proof needs to be slightly modified (as it appeals to the false Lemma 3.7) towards the end. The modification is as follows: we are in the situation of R_σ satisfying (ALG1), have chosen $a_1, \dots, a_n, a \in V(P)$ such that the field of definition of the ideal $P \cap R$ is contained in $k_0(a_1, \dots, a_n)$, and a is generic over $k_0(a_1, \dots, a_n)$. By (ALG1) and the way our coordinate rings are defined, we know that the ideal I of R_σ generated by $P \cap R$ is pure of dimension $\dim(P)$. As $V(I)$ has finitely many irreducible components and by genericity of a, a is in only one irreducible component of $V(I)$, and that component must be $V(P)$. Hence, for any ℓ, $P \cap R_{\sigma^\ell}$ is defined over $\mathrm{cl}_{\sigma^\ell}(k_0, a, a_1, \dots, a_n)$.

Corollary 4.2 and Propositions 4.3, 4.4 and 4.5 go through without change, except in the proof of 4.3, $(*1)$ should be replaced by the present Proposition 2.17.

In (4.6), we slightly strengthen the requirements and only consider 0-closed sets defined by *virtual perfect* ideals. This is to ensure that they have only finitely many irreducible components.

Proposition 4.7 remains true, with a slight change at the end of the proof, similar to the one given for 4.3.

Proposition 4.9 and Lemma 4.10 go through without change. Note the following consequences of Lemma 4.10 of [7], which while not needed for the main theorems, are quite useful in applications. We assume the hypotheses of 4.10.

Corollaries (of Lemma 4.10 of [7]). (1) *Let d_1 and d_2 be tuples of realizations of basic types among $\{p_1, \dots, p_n\}$. Then $\mathrm{acl}(d_1) \cap \mathrm{acl}(d_2) = \mathrm{acl}(e)$, where e consists of realizations of types in $\{p_1, \dots, p_n\}$.*

(2) *Let b realize a tuple of semibasic types, and $a \in \mathrm{acl}(b)$ be such that $\mathrm{qftp}(a/k_0)$ satisfies (ALGm) for some m. Then $\mathrm{qftp}(a/k_0)$ is semibasic.*

Proof. Choosing c in 4.10 to be the empty sequence, (1) follows from the conclusion.

(2) Indeed, without loss of generality b consists of realizations of basic types; take b' realizing $\mathrm{qftp}(b/a)$ and independent from b over a. Then $a = \mathrm{acl}(b) \cap \mathrm{acl}(b')$ and we may apply (1). $\square$

Let us now discuss Theorem 4.11. The set $\mathcal{Y}$ needs to be modified in the following manner:

- Condition (i) stays the same: for any semibasic type q, $\mathcal{X}_q(K) \subset \mathcal{Y}(K)$ or $\mathcal{X}_q(K) \cap \mathcal{Y}(K) = \varnothing$;
- Condition (ii) becomes: if $b \in \mathcal{Y}(K)^n$ for some n, and $a \in \mathrm{acl}(k_0 b)$ is such that $q = \mathrm{qftp}(a/k_0)$ satisfies (ALGm) for some m, then $\mathcal{X}_q(K) \subset \mathcal{Y}(K)$.

(The set $\mathcal{Y}$ was in fact incorrectly defined in [7], and the current definition is the one which was used in the proof.) In the cumulative case, we furthermore impose that all our semibasic types are cumulative.

Once this change is done, the proof goes through, although one needs to pay attention to a clash of notation: the tuple d which appears on line 13 of page 283 has nothing to do with the one discussed earlier in the proof; it consists of realizations of basic types, and is independent from c over k_0.

Proposition 4.12 of [7] goes through verbatim, as well as Remark 4.14, Proposition 4.15 and the verification of the axioms for Zariski geometries given in (4.16), for the set $\mathcal{Y}_b(K) = \bigcup_{p \text{ basic of dimension } e} \mathcal{X}_p(K)$. Note that the present Corollary 2.20 gives us Corollary 4.16 of [7] for semibasic types.

4. Using the Zariski geometry to get the trichotomy

The first paragraphs of Chapter 5 of [7] introduce Robinson theories and universal domains. The real work starts with Lemma 5.10 of [7], which out of a group configuration produces a quantifier-free definable subgroup of an algebraic group, in some reduct $\Omega[m]$. Note that in the cumulative case, the subgroup G_1 can be chosen so that its generic type is cumulative, by Proposition 1.15 of [5]. Then all results of [7] up to Proposition 5.14 go through without change.

Paragraph (5.15) is the statement of the trichotomy theorem:

Theorem 5.15. *Let p be a basic type, and assume that $\mathcal{X}_p(K)$ is not modular. Then $\mathcal{X}_p(K)$ interprets an algebraically closed field of rank* 1.

The proof given in [7] goes through, as it is just an adaptation of the proof of [11] to our particular case.

We now come to the main result of the paper, given at the beginning of Section 6:

Theorem. *Let $K \models$ ACFA, let $E = \mathrm{acl}_\sigma(E) \subseteq K$, and let p be a type over E with $\mathrm{SU}(p) = 1$. Then p is not modular if and only if p is nonorthogonal to the formula $\sigma^m(x) = x^{p^n}$ for some relatively prime $m, n \in \mathbb{Z}$ with $m \neq 0$.*

Proof. The proof goes through verbatim, to show that for some $m > 0$ (passing maybe to a larger E), if a realizes p, then there is some $a' \in \mathrm{acl}_\sigma(Ea)$ such that $\mathrm{evSU}(a'/E) = \mathrm{SU}(a'/E)[m] = 1$, and $\mathrm{qftp}(a'/E)[m]$ is nonorthogonal to the formula $(\sigma^m)^r(x) = \mathrm{Frob}^n(x)$ for some integers $r \neq 0$ and n, with $(n, r) = 1$ (and in fact, $r = 1$). The proof is now routine, using Lemma 1.12 of [2]: let b, c be tuples such that, in $\Omega[m]$, c is independent from $\mathrm{acl}_\sigma(Ea) = \mathrm{acl}_\sigma(Ea')$ over E, b satisfies $(\sigma^m)^r(x) = \mathrm{Frob}^n(x)$ and belongs to $E_0 = \mathrm{acl}_{\sigma^m}(Ea'c)$. The proof of Lemma 1.12 of [2] then gives us an $\mathrm{acl}_\sigma(Ea)$-σ^m-embedding φ of $F_0 = \mathrm{acl}_\sigma(Ea)E_0$ into $\Omega[m]$, such that the fields $\sigma^i\varphi(F_0)$, $i = 0, \ldots, m-1$, are linearly disjoint over $\mathrm{acl}_\sigma(Ea)$. It then follows that $\varphi(c)$ is independent from a over E (in Ω), and therefore p is nonorthogonal to $\sigma^{mr}(x) = \mathrm{Frob}^n(x)$. □

The proofs of the results of Section 7 are also unchanged.

We have proved one part of the trichotomy, namely the dichotomy between modularity and a field structure. The second leg is proved in all characteristics in (5.12) of [2]: if p is modular but has nontrivial algebraic closure geometry, then p is nonorthogonal to an SU-rank one definable subgroup of an algebraic group, indeed of the additive or multiplicative group, or a simple abelian variety.

Additional information concerning the nonorthogonality is available in [3; 4]. The internal structure of modular subgroups of semiabelian varieties is fully understood; see [6]. In the additive case, a bilinear map is definable in some cases; describing the full induced structure remains open.

References

[1] B. Castle, "Restricted Trichotomy in characteristic zero", preprint, 2022. arXiv 2209.00730

[2] Z. Chatzidakis and E. Hrushovski, "Model theory of difference fields", *Trans. Amer. Math. Soc.* **351**:8 (1999), 2997–3071. MR Zbl

[3] Z. Chatzidakis and E. Hrushovski, "Difference fields and descent in algebraic dynamics, I", *J. Inst. Math. Jussieu* **7**:4 (2008), 653–686. MR Zbl

[4] Z. Chatzidakis and E. Hrushovski, "Difference fields and descent in algebraic dynamics, II", *J. Inst. Math. Jussieu* **7**:4 (2008), 687–704. MR Zbl

[5] Z. Chatzidakis and E. Hrushovski, "An invariant for difference field extensions", *Ann. Fac. Sci. Toulouse Math.* (6) **21**:2 (2012), 217–234. MR Zbl

[6] Z. Chatzidakis and E. Hrushovski, "On subgroups of semi-abelian varieties defined by difference equations", *Trans. Amer. Math. Soc.* **369**:5 (2017), 3673–3705. MR Zbl

[7] Z. Chatzidakis, E. Hrushovski, and Y. Peterzil, "Model theory of difference fields, II: Periodic ideals and the trichotomy in all characteristics", *Proc. London Math. Soc.* (3) **85**:2 (2002), 257–311. MR Zbl

[8] R. M. Cohn, *Difference algebra*, Tracts Pure Appl. Math. **17**, Interscience Publishers, New York, 1965. MR Zbl

[9] B. Elsner, *Presmooth geometries*, Ph.D. thesis, University of Oxford, 2014, available at https://ora.ox.ac.uk/objects/ora:8878. MR

[10] E. Hrushovski, "The Manin–Mumford conjecture and the model theory of difference fields", *Ann. Pure Appl. Logic* **112**:1 (2001), 43–115. MR Zbl

[11] E. Hrushovski and B. Zilber, "Zariski geometries", *J. Amer. Math. Soc.* **9**:1 (1996), 1–56. MR Zbl

[12] Y. Peterzil and S. Starchenko, "Expansions of algebraically closed fields in o-minimal structures", *Selecta Math.* (*N.S.*) **7**:3 (2001), 409–445. MR Zbl

[13] T. Scanlon, "A positive characteristic Manin–Mumford theorem", *Compos. Math.* **141**:6 (2005), 1351–1364. MR Zbl

[14] O. Zariski and P. Samuel, *Commutative algebra*, vol. 1, Grad. Texts Math. **28**, Springer, 1975. MR Zbl

[15] B. Zilber, "A curve and its abstract Jacobian", *Int. Math. Res. Not.* **2014**:5 (2014), 1425–1439. MR Zbl

Received 29 Dec 2022. Revised 18 Feb 2024.

ZOÉ CHATZIDAKIS:

zoe.chatzidakis@imj-prg.fr
Université Paris Cité and Sorbonne Université, CNRS, IMJ-PRG, F-75013 Paris, France

EHUD HRUSHOVSKI:

ehud.hrushovski@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.305

Boris Zilber and the model-theoretic sublime

Juliette Kennedy

We examine some of Zilber's early theorems through the lens of the "model-theoretic sublime".

1. Introduction

A recent email exchange between myself and Boris Zilber, stimulated by a lecture he gave in Helsinki a few years prior, began by discussing his various moves to generalize the syntax/semantics distinction.[1] Boris's repurposing of the syntax/semantics distinction — a distinction taken more or less for granted in foundational practice — has always been interesting; but in our exchange Boris also broke out *philosophically*:

BZ: These are, I guess, two ways of how we perceive the world: the intellectual, words-based way, and the intuitive, sensory way. In mathematics, the first way requires you to write down a full proof of the fact (the ultimate explanation). The second, semantical way, is to see a picture, mental or graphical, that talks to your experience of the world. It is also what is responsible [for the] division of mathematics into Algebra and Geometry. Michael Atiyah (in his millennium lecture?) says that Geometry-Algebra is like Space-Time pairing: In geometry you see the whole at once, no time needed. In algebra you need time to read it letter-by-letter, but not space.

The words-based way and the semantical way, to wit: the mathematician is tethered to the *sign*, to formal correctness and to the "letter-by-letter" of proof; while on the other hand there is *insight* and *experience*, *meaning* and *seeing the whole picture*. Two poles pulling away from each other, and the mathematician caught somewhere in between.

I would like to thank the Academy of Finland for their support, grant no. 322488. This project has also received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 101020762). I am also grateful to Jonathan Kirby and to two anonymous referees for their helpful comments.

MSC2020: primary 00A30; secondary 03C45.

Keywords: sublime, Kant, trichotomy, surveyability, Zilber, Pallasmaa.

[1] The email exchange took place in May of 2023.

In this note I would like to think about the way Boris pulls the curtain back on this binary practice of the mathematician, in his rich remark to me, so full of philosophical moves. One is struck by the phrase “seeing the whole at once, no time needed” — a move toward the *sublime*, I suggest, an aesthetic category important in 18th century British philosophy and of renewed interest today in the form of, for example, the environmental sublime.[2]

Thinking through Boris’s beautiful remark in the context of the sublime helps us to place his remark, and beyond that his mathematical work, *philosophically*. Generally speaking, the philosophical content of a foundational attitude often has to do with its (so-called) existential or metaphysical commitments — or its lack thereof: how entangled with set theory it is, its putative second-order content, the theory’s constructive content, and so on. I would like to think about Boris’s work, though, by drawing on ideas coming from somewhat outside the foundations of mathematics culture. One is, of course, not *against* foundations of mathematics; for, to paraphrase Emily Apter [2013, p. 2], if one is against foundations, as a logician, what could one possibly be for? It is just that the interest here is in developing novel interpretive strategies.

2. The sublime

Boris’s phrase “seeing the whole at once, no time needed” reminded me of the remark of the 18th century aesthetic theorist Alexander Gerard, that sublimity is the state in which “the mind . . . imagines itself present in every part of the scene it contemplates” [Gerard 1759, p. 14].[3]

More commonly[4] the sublime is thought of in the terms Kant laid out for it, namely in terms of a physical immensity, usually in nature — think of standing at the precipice of an enormous crevasse — that pitches the subject into a kind of vertigo; “doing violence to the imagination”, as Kant put it; leaving the subject’s cognitive apparatus undone. As Emily Brady writes, on the Kantian sublime:

> The sources of the sublime response are linked to the physical properties of magnitude or power in nature but importantly also to the failure of imagination, without which it could not occur. Imagination’s activity in the sublime, in contrast to the beautiful, is “serious”, where some object is “contrapurposive for our power of judgment, unsuitable for our faculty

[2] See, e.g., [Brady 2013]. Kant discusses the sublime mainly in the so-called third critique, the *Critique of the power of judgment* [Kant 2000, Sections 25–28].

[3] Gerard continues: “. . . and from the sense of this immensity, feels a noble pride, and entertains a lofty conception of its own capacity.”

[4] More commonly in the philosophical aesthetics literature at least. As a referee has helpfully pointed out, the *everyday* meaning of the term “sublime”, evoking properties such as “calmness” or “beauty”, differs markedly from its meaning in philosophy.

> of presentation, and as it were doing violence to our imagination, but is nevertheless judged all the more sublime for that."[5]

This "astonishment bordering on terror", as Kant rather hyperbolically called it, involves anxiety, then, bordering on fear, but also, somehow, pleasure: the pleasure of being in the vicinity of danger, while at the same time being out of it; the pleasure of being in awe of something. *Negative* pleasure was Kant's term for this, while positive pleasure is pleasure in the beautiful, which "brings with it a promotion of life" [Kant 2000, p. 128].[6] Interestingly enough, because pleasure is involved, sublimity is theorized by philosophers an *aesthetic* category. The sublime response, in other words, is an *aesthetic* response.

Kant distinguishes the dynamic sublime, in which the subject is undone, so to speak, by a natural scene, from the mathematical sublime, in which the subject experiences a failure of the imagination, not in the face of a natural immensity but in the face of an infinite number sequence. In the encounter with the mathematical sublime the subject is thrown into confusion once again, for not having a grip on the contours of the thing at hand; but also being inexorably compelled. One might call this mixture of attraction and unease the *mathematician's* negative pleasure.

Kant's observation was that although the senses fail to deliver a conceptual unity on their own, the sequence can nevertheless "be *completely* comprehended *under* one concept", and this is due to a faculty of "suprasensible" reason:

> And what is most important is that to be able only to think it [the infinite: JK] as a *whole* indicates a faculty of mind which surpasses every standard of sense .... Nevertheless, *the bare capability of thinking* this infinite without contradiction requires in the human mind a faculty itself suprasensible.[7]

In other words, Kant gives us what the (classical) mathematician would say is the correct outcome. Reason meets the imagination at its point of collapse, delivering the infinite object as a conceptual unity — just as reason delivers a phenomenal unity in the case of overwhelming natural phenomena. The important point here is that this faculty of reason is *suprasensible*, so going beyond (sense) experience.

Sublimity, then, is not an experience of defeat, or not wholly; one is able to move out of it with the help of the mind's ability to synthesize the atomized scene, to structure chaos as nonchaos[8] — by means of a conception.

---

[5] See [Brady 2013, p. 57]. The interior quote is from Kant's *Critique of the power of judgment* [Kant 2000, p. 129].

[6] See also [Brady 2013, p. 57].

[7] [Kant 2000, Section 26, p. 94]. See also [Ginsborg 2005].

[8] The expression is due to the artist Eva Hesse [Bourdon 1970].

As an aside, the sublime has a moral dimension, putting us in touch, as Brady [Brady 2013, p. 59] writes, with our moral capacities. The sublime tutors us in "[loving] something, even nature, without interest . . . even contrary to our (sensible) interest" [Kant 2000, p. 151]. Witnessing the failure of the imagination, the failure of her imagination to comprehend the scene, the subject remains "undemeaned", as Kant put it, even so, and even has a feeling of superiority over nature, or in our case, the mathematical field, while at the same time "the human being must submit to that dominion" [Kant 2000, pp. 261–262]. A century later Leo Marx would coin the term "the *technological* sublime" to describe the conflict arising from holding the romantic (sublime) conception of the American landscape of the late 19th century, seeing that terrain as a kind of virginal paradise, while employing the rhetoric of industrial progress [Marx 1964, p. 7]. And just a few years after that Hilbert would lace his oft-cited 1930 "ignorabimus" address with the language of human supremacy, expressed in terms of the technological optimism typical of the period.

Sublimity, in other words, is always connected to *power*. In the wake of the various emergencies, climatic and otherwise, besetting human beings in the 21st century, it is not surprising that there is a renewed philosophical interest in the sublime!

Later, post-Kantian and post-Gerardian passes at the sublime by writers such as F. R. Ankersmit would untether sublimity from awe and the idealization of nature that was characteristic of the earlier theories, so that the sublime could now be deployed in other domains, such as history, or psychoanalysis.[9] Ankersmit in particular took a *melancholic* view of sublimity, emphasizing the static quality of

[9]From [Ankersmit 2005, p. 335]:

> The traumatic experience is too terrible to be admitted to consciousness: The experience exceeds, so to speak, our capacities to make sense of experience. Whereas normally the powers of association enable us to integrate experience into the story of our lives, the traumatic experience remains dissociated from our life's narrative since these powers of association are helpless and characteristically insufficient in the case of trauma. And there is one more resemblance between trauma and the sublime that is of relevance in the present context. Characteristic of trauma is the incapacity to actually suffer from the traumatic experience itself . . . . The subject of a traumatic experience is peculiarly numbed by it; he is, so to speak, put at a distance from what caused it. The traumatic experience is dissociated from one's "normal" experience of the world . . . . Now, much the same can be observed for the sublime. When Burke speaks about this "tranquility tinged with terror," this tranquility is possible (as Burke emphasizes) thanks to our awareness that we are not really in danger. Hence, we have distanced ourselves from a situation of real danger — and in this way, we have dissociated ourselves from the object of experience. The sublime thus provokes a movement of derealization by which reality is robbed of its threatening potentialities. As such Burke's description of the sublime is less the pleasant thrill that is often associated, with it than a preemptive strike against the terrible.

the sublime response, the idea that the subject is locked into a back-and-forth cycle of attraction and repulsion. Sublimity, in other words, is a site of conflict:

> Now, aesthetics provides us with the category of the sublime for conceptualizing such a conflict of schemes without reconciliation or transcendence. Thus the Kantian sublime is not a transcendence of reason and understanding and the entry to a new and higher order reality, but can only be defined in terms of the inadequacy of both reason and understanding .... Similarly, it is only by way of the positive numbers that we can get access to the realm of negative numbers; and gaining this access does not in the least imply the abolition or transcendence of the realm of the positive numbers, but a continuous awareness of their existence as well.[10]

Kant's account of the role of intuition and reason in delivering conceptual coherence within sublimity is embedded in a complex theory of the mind, one drawing on specific conceptualizations of the faculties of imagination and reason. Kant's theory of the mathematical sublime is about our mathematical capacities überhaupt, and as such it slots easily into the contemporary conversation in the foundations of mathematics, the debates about the nature of finitary intuition, or what constitutes a genuinely constructive proof.

What rather holds my interest in thinking through Boris's beautiful remark though, are not the foundational issues per se, but the way his remark reveals that logic too is a site of conflict: a conflict that gets read into the syntax/semantics distinction, a conflict that renders logic so alive philosophically. It is astonishing that logic can even take the exact mathematical measure of that conflict, that is to say drawing out deep theorems from it, limitative results such as the incompleteness theorems due to Gödel, or the undefinability of truth due to Tarski.

In Ankersmit's writings the mathematical sublime is domesticated, as it were, so that mathematical sublimity now signifies anything in the way of a mathematical unknown:

> Think of the equation $f(x) = \frac{1}{3}x^3 + \frac{1}{2}x^2 - 12x$. Differential calculus shows that this function will have a local maximum for $x = -4$ and a local minimum for $x = 3$. In this way differential calculus can be said to perform what, analogously, could not possibly be performed for the relationship between narrative and experience. So one might say that historical writing is in much the same situation as mathematics was before the discovery of differential calculus by Newton and Leibniz. Before this discovery there was something "sublime" about the question of where the equation $f(x) = \frac{1}{3}x^3 + \frac{1}{2}x^2 - 12x$ would attain its local optimum and minimum: One could only hit on it experimentally (that is, by simply

[10] [Ankersmit 2002, p. 207].

> trying out different values for x), but no adequate explanation could be given for this. It has been Newton's and Leibniz's feat of genius to reduce what was "sublime" to what could be figured out, or to reduce what was incommensurable to what could be made commensurable thanks to the magic of differential calculus.[11]

Ankersmit is thinking about sublime historical experience in this passage, but we can draw the moral from it that Kant's notion of the mathematical sublime (which applied only to extended objects) was too narrow. It is not just that the imagination cannot take in infinite totalities; the mathematical field is full of concepts and ideas that cause the mathematician to lose his footing. There is the concept of a model class, for example — or how to get a foothold there? In logic one has the space of all first-order theories — how to find a way through that morass? Set theory also, with its large cardinal hierarchy, is threaded with sublimity through and through.

3. Categoricity and classification

Let us now turn to Boris's work, in particular its synthetic aspect within what one might call the *model-theoretic sublime*. Let us take "synthesis" to refer to an act of (mathematical) reason that structures some heretofore unstructured part of the mathematical field — unstructured in the sense of being untheorized, or unclassified, or simply formless.

The suggestion here is that both categoricity and classification can be viewed as devices imposing structure on the mathematical field, albeit in different ways: categoricity, a notion occupying a central place in Boris's mathematical work, by collapsing the space of all possible models (of a fixed cardinality) of an uncountably categorical theory to a single point (up to isomorphism);[12] classifiability in virtue of being an organizing principle, a kind of scaffolding structure for the space of first-order theories.[13]

Categorical theories are "logically perfect", in Boris's terminology, where logical perfection means the following: "... a mathematical object of a certain 'size' is logically perfect if in a certain formal language it allows a 'concise' description fully determining the object" [Cruz Morales et al. 2021, p. 2]. Precisely:

> The amazing conclusion derived from the research is that among the huge diversity of mathematical structures there are very few which satisfy the (slightly narrower) definition of categoricity, and those can be classified.

[11] [Ankersmit 2005, p. 175]. Ankersmit's reading of the historical details with regard to commensurability may be regarded by some as contentious.

[12] A theory T is said to be "categorical" if T has a unique model, up to isomorphism. T is said to be "categorical in power" if for all cardinals κ, T has a unique model of size κ, up to isomorphism.

[13] For an example of classifiability, see the below discussion of the main gap theorem.

> These certainly seem to corresponding to an ideal of logical perfection, in the following sense: categorical structures M determine a first-order theory $\mathrm{Th}(M)$ (the set of all sentences that are true in M) and then comes the reason why we call them "logically perfect": all other structures that satisfy the theory $\mathrm{Th}(M)$ and are of the same cardinality as M are isomorphic to M. In other words, uncountably categorical structures are inextricably linked to their logical description; the description $T = \mathrm{Th}(M)$ completely determines the structure M.[14]

The search for categorical axiomatisations of canonical mathematical theories is a philosophical project, fundamentally, albeit one pursued entirely within mathematics (or, precisely, within mathematical logic). If our canonical mathematical theories have a unique interpretation, referential indiscernibility is eliminated — which is just simply to say that in mathematics, or at the very least in the case at hand, we really do "mean what we say". For that reason, perhaps, categoricity represents, for Boris, the apotheosis of logical perfection. He has even conjectured, boldly, that "Categoricity is bound to play the role that analyticity played for number theory, but for physics" (see [Villaveces 2022]).

Andrés Villaveces writes eloquently about the epistemological aspect of categoricity, its evidentiary force, in a remark that seems, somehow, to gesture at sublimity:

> Al enfrentarnos a ciertas descripciones o afirmaciones nuestra reacción natural de incredulidad puede ser vista como una de las raíces de la búsqueda de atrapar, aprehender, mediante el lenguaje, la descripción de un fenómeno, de un objeto matemático o de un evento. Al vernos enfrentados a una afirmación (matemática o no), la primera reacción natural en muchas circunstancias suele ser de incredulidad. Ante la duda, intentamos buscar confirmación a como dé lugar. Dejando de lado búsquedas de verificación por autoridad, podemos señalar dos grandes tipos de confirmación: por verificación directa, por una buena descripción de la teoría que sustenta la afirmación en cuestión.[15]

---

[14] [Cruz Morales et al. 2021, p. 6].

[15] [Villaveces 2022]. In translation:

> When faced with certain descriptions or statements, our natural reaction of disbelief can be seen as one of the roots of the search to capture, apprehend, through language, the description of a phenomenon, of a mathematical object. or an event. When faced with a statement (mathematical or not), the first natural reaction in many circumstances is usually disbelief. When in doubt, we try to seek confirmation no matter wherefrom. Leaving aside verification by authority, we can point out two main types of confirmation: by direct verification, or by a good [i.e., categorical: JK] description of the theory that supports the statement in question.

Categorical theories are “logically perfect”, in Boris’s terminology, not only because they provide a compact description of a seemingly intractable field of concepts, but for enabling the possibility of regarding space as a coherent way of pasting localized versions of itself — a perfection realized, for Boris and coauthors in [Cruz Morales et al. 2021], by the notion of an affine scheme due to Grothendieck.

Synthesis emerges in model theory also through classification. Instead of a heterogeneous collection of theories (so theories this time, instead of models), and the mathematician having to creep from one theory to the next, to paraphrase Gerard,[16] Boris offered up the trichotomy conjecture,[17] which turned out to hold of the (very ample) Zariski structures:

Conjecture. If X is a strongly minimal set, then exactly one of the following is true about X.

(1) X is trivial in the sense that algebraic closure (on a saturated model of the theory of X) defines a degenerate pregeometry (for any set $A \subseteq X$ one has $\mathrm{acl}(A) = \bigcup\{\mathrm{acl}(\{a\}) \mid a \in A\}$).

(2) X is essentially a vector space. That is, possibly after adding some constant symbols to the language of X, there is an infinite group space G bi-interpretable with X for which every definable subset of any Cartesian power of G is a finite Boolean combination of cosets of definable subgroups.

(3) X is bi-interpretable with an algebraically closed field.[18]

Classification theorems in mathematics, then, serve as a move toward synthesis: *resisting* or *dissolving* sublimity, structuring the heretofore unstructured mathematical field as nonchaos, providing a scaffolding.

Together with Boris’s work on trichotomy one should mention Shelah’s main gap theorem [Shelah 1990], which is another masterpiece in the genre of classification theorems. The theorem states that the class of all first-order theories falls into two categories: the tame or classifiable, and the nonclassifiable. The former have “few” models and admit a dimension-like set of geometric invariants; the latter have the

[16] Gerard [1759] remarks:

> Objects cannot possess that largeness, which is necessary for inspiring a sensation of the sublime, without simplicity. Where this is wanting, the mind contemplates, not one large, but many small objects: it is pained with the labour requisite to creep from one to another; and is disgusted with the imperfection of the idea, with which, even after all this toil, it must remain contented. But we take in, with ease, one entire conception of a simple object, however large: in consequence of this facility, we naturally account it one . . . the view of any single part suggests the whole, and enables fancy to extend and enlarge it to infinity, that it may fill the capacity of the mind.

[17] See [Zilber 1984]. For Hrushovski’s counterexample, see [Hrushovski 1993]. For the trichotomy theorem, see [Hrushovski and Zilber 1993].

[18] For a survey of recent work in the area see [Baldwin and Villaveces 2024] in this volume.

maximum number of models possible, and are entangled with each other in a way that makes it difficult to tell some of them apart.[19] The main gap theorem almost seems to be written in the language of the sublime!

4. Geometry as place

Returning to Boris's philosophical remark, if "in geometry you see the whole at once, no time needed", one may ask, what is this *"whole"* that Boris sees at once, no time needed? I would like to touch down here, albeit lightly, in the notion of *place*. Perhaps what geometry allows one to see is a kind of *place* — not in any literal sense but in the sense that the architectural theorist Juhani Pallasmaa means in his writings about placeness: a site of experiential cohesion, one resonating "with the inner qualities of placeness in our minds ... a constitutive condition for anything to exist in human consciousness" [Pallasmaa 2023]. Pallasmaa states:

> The experience of placeness can ... arise from countless characteristics and features, but fundamentally it is a consequence of experiential cohesion, spatial or formal singularity, communal agreement, or meaningfulness of a distinct entity in the physical world .... Through constructions, both material and mental, useful and poetic, practical and metaphysical, we create places, existential footholds in the otherwise meaningless world.

The thought here is that through geometry and its suggestion of place, through thinking of geometry as creating the conditions for the notion of place, the mathematician is led toward the possibility of concretizing, structuring, contextualizing and internalizing mathematical ideas. Note that we take places in at once, no time needed. As Pallasmaa puts it, "We 'understand' qualities of places unconsciously before we have had any chance for intellectual evaluation or understanding."

If architecture, for Pallasmaa, is engaged with the lived meaning of space, "[projecting] predictable order and meaning into human existence", "[mediating] between the threatening immensity of the world, the infinity and anonymity of space, as well as the endlessness of time", here geometry stands in for, in the sense of functioning as, architecture, in grounding the mathematician in the mathematical field, in enabling the possibility of lived mathematical experience.

[19] More precisely, Shelah's main gap theorem divides all countable, complete first-order theories into two categories: in the classifiable case, there is a bound on the number of models (up to isomorphism), and they can be characterized by a tree of geometric invariants, like the dimension of a vector space, while at the same time in the nonclassifiable case, there is a precise sense in which no notion of dimension can be extracted, and the case is chaotic in the sense that the structures are hard to tell apart.

There is also the ontological question: is anything real in mathematics, that is *not* related to geometry? “Nothing is that is not placed”, as Plato has reportedly said.[20]

5. Conclusion

Amid the debates in philosophical aesthetics, such as whether aesthetic properties reside in the subject or in the object, or whether aesthetic experience involves cognition or not, the sublime persists as a central irritant. In the hands of contemporary philosophers the sublime has been extended well beyond the categories Kant envisaged, as we saw, so that we now have the romantic sublime, the technological sublime, the environmental sublime that Emily Brady writes about so eloquently, the historical sublime, the moral sublime, and so on. There is a substantial philosophical literature, by now, on the sublime; let us add to it the category of the *model-theoretic sublime*.

In this brief note I have strayed into philosophical territory; but in fact the correspondence between Boris and I ended with Boris going to ground philosophically:

JK: In your own work though, how is it helpful to think of the syntax/semantics distinction in the way you do?

BZ: …here is one of my talks on the topic, attached. It is what resulted from my attempts to understand what “non-commutative geometry” is and how it originated in Heisenberg’s physics. In more detail, you can download a couple of papers from my web-page, like “The geometric semantics of algebraic quantum mechanics”.

Boris’s mathematical work stages a beautiful encounter with the mathematical sublime. It is essential that we recognize it as a logician’s encounter with the mathematical sublime, that is to say, one occurring within *logic*. This is because the display of power here originates exactly in the logician’s gift, unique to him among all mathematicians, namely his sensitivity to language — utilizing and directing that power, as Boris does, onto mathematics and physics.

In writing model theory in the language of geometry, a hallmark of Boris’s mathematical practice, the conflicted aspect of sublimity, the idea of stasis and being locked into a cycle, is set aside, and the conditions for a *rapprochement* between the words-based way and the semantical way are laid out, because of geometry. It is a road that opens out into freedom for the logician; it is a road that delivers the logician into the mathematical arena. And while coming to grips with Boris’s work involves a great deal of negative pleasure — because of the toil

[20] Jeff Malpas in his lecture at the *Understanding and Designing Place* Symposium at the Tampere University on 3 April, 2017.

involved but also being, as we are, in awe of what he has done — then if positive pleasure is pleasure in the *beautiful*, simply and for itself, Boris's work gives us that too — straight to the heart.

References

[Ankersmit 2002] F. R. Ankersmit, *Historical representation*, Stanford Univ. Press, 2002.

[Ankersmit 2005] F. R. Ankersmit, *Sublime historical experience*, Stanford Univ. Press, 2005.

[Apter 2013] E. Apter, *Against world literature*: *on the politics of untranslatability*, Verso, Brooklyn, NY, 2013.

[Baldwin and Villaveces 2024] J. T. Baldwin and A. Villaveces, "Zilber's notion of logically perfect structure: universal covers", *Model Theory* **3**:2 (2024), 647–683.

[Bourdon 1970] D. Bourdon, "Fling, dribble and drip", *Life* **68**:7 (1970), 62–67.

[Brady 2013] E. Brady, *The sublime in modern philosophy: aesthetics, ethics, and nature*, Cambridge Univ. Press, 2013.

[Cruz Morales et al. 2021] J. A. Cruz Morales, A. Villaveces, and B. Zilber, "Around logical perfection", *Theoria* **87**:4 (2021), 971–985. MR Zbl

[Gerard 1759] A. Gerard, *An essay on taste*, London, 1759.

[Ginsborg 2005] H. Ginsborg, "Kant's aesthetics and teleology", in *The Stanford encyclopedia of philosophy*, edited by E. N. Zalta, 2005.

[Hrushovski 1993] E. Hrushovski, "A new strongly minimal set", *Ann. Pure Appl. Logic* **62**:2 (1993), 147–166. MR Zbl

[Hrushovski and Zilber 1993] E. Hrushovski and B. Zilber, "Zariski geometries", *Bull. Amer. Math. Soc.* (*N.S.*) **28**:2 (1993), 315–323. MR Zbl

[Kant 2000] I. Kant, *Critique of the power of judgment*, Cambridge Univ. Press, 2000.

[Marx 1964] L. Marx, *The machine in the garden*: *technology and the pastoral ideal in America*, Oxford Univ. Press, 1964.

[Pallasmaa 2023] J. Pallasmaa, "From space to place — existential meaning in architecture", *Tangible Territory* **1**:5 (2023).

[Shelah 1990] S. Shelah, *Classification theory and the number of nonisomorphic models*, 2nd ed., Studies in Logic and the Foundations of Mathematics **92**, North-Holland Publishing Co., Amsterdam, 1990. MR Zbl

[Villaveces 2022] A. Villaveces, "Atrapar suavemente y dejar ir: categoricidad y lugar", pp. 83–98 in *Rastrear indicios*, Textos: Documentos de Historia y Teoría **29**, Editorial Universidad Nacional de Colombia, 2022.

[Zilber 1984] B. I. Zilber, "Strongly minimal countably categorical theories, II", *Sibirsk. Mat. Zh.* **25**:3 (1984), 71–88. In Russian; translated in *Sib. Math. J.* **25** (1984), 396–412. MR Zbl

Received 20 Sep 2023. Revised 24 Nov 2023.

JULIETTE KENNEDY:

juliette.kennedy@helsinki.fi
Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.317

Approximate equivalence relations

Ehud Hrushovski

This paper is dedicated to Boris Zilber, who found the path for us.

Generalizing results for approximate subgroups, we study approximate equivalence relations up to commensurability, in the presence of a definable measure.

As a basic framework, we give a presentation of probability logic based on continuous logic. Hoover's normal form is valid here; if one begins with a discrete logic structure, it reduces arbitrary formulas of probability logic to correlations between quantifier-free formulas. We completely classify binary correlations in terms of the Kim–Pillay space, leading to strong results on the interpretative power of pure probability logic over a binary language. Assuming higher amalgamation of independent types, we prove a higher stationarity statement for such correlations.

We also give a short model-theoretic proof of a categoricity theorem for continuous logic structures with a measure of full support, generalizing theorems of Gromov–Vershik and Keisler, and often providing a canonical model for a complete pure probability logic theory. These results also apply to local probability logic, providing in particular a canonical model for a local pure probability logic theory with a unique 1-type and geodesic metric.

For sequences of approximate equivalence relations with an "approximately unique" probability logic 1-type, we obtain a structure theorem generalizing the "Lie model" theorem for approximate subgroups (Theorem 5.5). The models here are Riemannian homogeneous spaces, fibered over a locally finite graph.

Specializing to definable graphs over finite fields, we show that after a finite partition, a definable binary relation converges in finitely many self-compositions to an equivalence relation of geometric origin. This generalizes the main lemma for strong approximation of groups.

For NIP theories, pursuing a question of Pillay's, we prove an archimedean finite-dimensionality statement for the automorphism groups of definable measures, acting on a given type of definable sets. This can be seen as an archimedean analogue of results of Macpherson and Tent on NIP profinite groups.

Research funded in part by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement No. 291111.
MSC2020: 03C07.
Keywords: approximate equivalence relations, NIP, probability logic.

1. Introduction

Let G be a group, with a translation-invariant, finitely additive measure on some Boolean algebra of subsets of G. For instance G may be an ultraproduct of finite groups with their counting measures. A symmetric subset X of a group G is called a *near-subgroup*[1] if both X and the triple product set XXX are measurable and of finite, nonzero volume. These are very closely connected to amenable *approximate subgroups*, and arise in many branches of mathematics; see, e.g., [17] for an introduction.

If X is a Lie group, such as $\mathrm{GL}_n(\mathbb{R})$, a compact neighborhood of the identity is a near-subgroup; all such neighborhoods are commensurable, i.e., each is covered by finitely many translates of the other. Conversely it was shown in [43] that a near-subgroup determines canonically a connected Lie group L, so that X is commensurable with a pullback of a compact neighborhood of the identity in G.

This was used in particular to give a proof of (a strengthening of) Gromov's polynomial growth theorem, based on measure-theoretic rather than metric properties of such a group.

Gromov was at that time writing [35]. He wrote:[2]

> I think there are many "almost structures" which are far from "actual structures" and these may play an essential role in how the brain generates math (...) the "dictionary structure" contains algebraic pattern, e.g., of "categories", "multicategories" and also of "2-categories" but these patterns are "not perfect", e.g., some compositions may be "undefined/not implemented" and some may be nonassociative.

This comment precipitated the current work. The simplest kind of category is a groupoid, and the simplest groupoids, other than groups, are equivalence relations. Transposing the result on statistical recognition of approximate subgroups to approximate equivalence relations thus appears as a natural first step. An equivalence relation is a symmetric relation R satisfying $R \circ R = R$; an approximate equivalence relation is one where $R \circ R$ is "commensurable" to R, i.e., all the $R \circ R$-neighbors of any point a are R-neighbors of some finite number of points; see Definition 4.1. In case X is an approximate subgroup of a group G, one can define $R(x, y) := x^{-1}y \in X$; then it is easy to see that R is an approximate equivalence relation. It will turn out in fact that using this construction, the results we will obtain on approximate equivalence relations imply the ones on approximate groups; the latter appears as the special case of approximate equivalence relations with a transitive automorphism group.

[1] The definition of near-subgroup in [43] is a little more general.

[2] Personal communications to the author, 15 and 17 September 2010.

The language of the paper is that of probability logic; we will give details of that separately below.

Stabilizer theorem. The dimension-theoretic stabilizer was first introduced to model theory by Boris Zilber, in the setting of groups of finite Morley rank. It was transformative to the subject, enabling for instance the proof of Zilber's indecomposability theorem, that seemed previously to belong to geometry rather than model theory. The construction was generalized to stable theories; the uniqueness of independent pairs, given their individual types, was key. Later (see [40; 24; 55]) it was realized that a good theory of the stabilizer exists without this uniqueness, under the weaker condition of the *independence theorem*. This statement asserts the existence, under suitable conditions, of a 3-type with three prescribed restrictions to independent 2-types.

Still later, it was possible to transpose these ideas to groups defined in any theory, carrying a suitable measure. This was the basis of a connection between near subgroups and the locally compact world: for near subgroups X, the stabilizer is a $\bigwedge$-definable group S contained in X^4, such that X/S is compact. Equivalently, one can find definable sets Y commensurable to X allowing a prescribed number of multiplications staying within X^4.

In Section 4 we will prove a generalization of the stabilizer theorem to near equivalence relations. Theorem 4.3 provides a canonical $\bigwedge$-definable equivalence relation S contained in $R^{\circ 4}$, such that each neighbor set of R is compact modulo S.

Riemannian homogeneous spaces. In the case of groups, one can go further and describe approximate subgroups up to commensurability as approximations to finite-dimensional Lie groups. In Section 5 we obtain a similar theorem under an assumption of *approximate homogeneity*. The model spaces are now Riemannian homogeneous spaces, with a "mesoscopic" graph relation connecting two points at distance at most 1. These are fibered over locally finite graphs; we obtain only a partial description of the total space, but a full description of each connected component. See Theorem 5.5.

Pseudofinite fields. Section 6, "From groups to graphs", concerns approximate equivalence relations definable in pseudofinite fields.

In [44], a model-theoretic proof was given of the "strong approximation" theorem of Gabber, Matthews–Vaserstein–Weisfeiler, Gabber and Nori (see [66; 67]) on subgroups of algebraic groups over the p-element field $\mathbb{F}_p$ for large p. The main model-theoretic ingredient was the study of generation of groups by definable sets. This is generalized in Section 6 to the generation of equivalence relations by definable relations. An arbitrary definable relation is decomposed, in each piece of a partition, into relations that generate an equivalence relation in finitely many steps, and relations of finite valency. For pseudofinite fields this decomposition has an explicit algebraic form; but the general result is proved in the setting of simple theories with a well-behaved finite dimension theory.

The measure stabilizer in NIP theories. In Section 7 we take an alternate route to finite-dimensionality, under an assumption of NIP. The passage from locally compact spaces to finite-dimensional Riemannian manifolds in Section 5 involved factoring out a large compact normal subgroup; commensurability is preserved, but little control over this compact kernel is available. In particular for a near-subgroup X of a compact group G, i.e., a definable subset of positive measure, this procedure loses all information. In Section 7 we factor out only the measure-theoretic stabilizer of X, and prove, assuming NIP, that up to possible profinite parts, the result is a finite-dimensional Lie group.

More generally, let μ be a definable measure on X, let $q(u)$ be a type and let $\phi(x, u)$ be a NIP formula. The formula ϕ establishes a relation between the space X of weakly random global types on X and the space of Kim–Pillay strong types extending q. We obtain corresponding quotients $\mathsf{X}_{\phi,q}$ of X, and a canonical space $\mathsf{U}_{\mu,\phi}$ of strong types compatible with q. The automorphism group of any saturated model induces a compact group $G_{\mu,\phi,q}$ acting faithfully on both $\mathsf{X}_{\phi,q}$ and $\mathsf{U}_{\mu,\phi}$. We prove that $G = G_{\mu,\phi,q}$ has finite archimedean rank. This means that G has a minimal normal subgroup G^{00} with G/G^{00} profinite, a maximal (up to finite index) profinite normal subgroup G_{00}, and G^{00}/G_{00} is a finite-dimensional Lie group.

This result can be transposed from automorphism groups to definable groups. Let G now be a definable group in a NIP theory, and μ a translation invariant definable measure on G. G has a minimal $\bigwedge$-definable subgroup G^{00}, and $K = G/G^{00}$ is compact in the logic topology. These compact groups were the subject of Pillay's conjectures in the o-minimal case (see [68; 47]) showing that K is a Lie group of the same dimension as G. In the general NIP case, beyond compactness, the constraints on K are unclear. However one can define a canonical quotient K_P of K associated with a given definable subset P of G, at least when G carries a definable measure μ; namely, identify two weakly random types of G if they include the same set of translates of P. (If $P = P_b$ is defined only with a parameter b, identify p with p' if for any $b' \models \mathrm{tp}(b)$ and any $g, g' \in G$, $g' P_{b'} \in p$ iff $g P_b \in G$.) It was also Pillay who had the intuition that K_P may be finite-dimensional. We show indeed that K_P has finite archimedean rank (Theorem 7.15). In fact this is what motivated the more general Theorem 7.12.

In view of results of Macpherson and Tent, it seems possible that a similar finite-dimensionality phenomenon is valid for the totally disconnected part of G, and in fact G is of adelic origin; in the most optimistic scenario, G is interpretable in the model-theoretic sum of finitely many p-adic and real fields. A generalization to the setting of approximate groups (viewed as piecewise-definable groups), using Gleason–Yamabe theory in place of Peter–Weyl, would also be interesting. See Section 7.18.

Probability logic. In probability logic, existential and universal quantifiers are augmented by *probability quantifiers*. If as in [53] they are entirely replaced by such quantifiers, we refer to *pure probability logic*. Events are taken to be definable subsets of a sort X, declared to be a *stochastic sort*. $E_x\phi(x, y)$ is a formula with free variables y, giving at $y = b$ the probability of the event $e_{\phi,b} = \{x : \phi(x, b)\}$.

Since $E_x\phi(x, y)$ takes real values, it is natural to allow all formulas to take bounded real values and employ continuous logic. When ϕ takes values other than 0 and 1, $E_x\phi(b)$is understood as the *expectation* of $\phi(x, b)$.

The axioms, going back to Kolmogorov, S. Bernstein, R. von Mises, Hilbert and Bohlmann, are just finite additivity and positivity. Countable additivity is not assumed but is automatically obtained for the induced measure on the type spaces over a model.

By iterating the expectation quantifiers, we obtain measures on type spaces in several variables too. The action of the symmetric group $\mathrm{Sym}(n)$ on the space of n types is not assumed to preserve the measure; when it does, we say that Fubini holds. Note that Fubini's *theorem* relates to the algebra generated by rectangles; the Fubini *property* goes beyond this to arbitrary binary relations. We do not assume Fubini at the level of the definition, but many results have stronger versions if Fubini is assumed.

We give a simple model-theoretic proof of a uniqueness theorem for models where every open set has positive measure. This generalizes a theorem of Keisler's on uniqueness of hyperfinite models, and theorems of Gromov and Vershik on invariants for measured metric spaces. When it exists the full-support model provides a canonical, homogeneous model for pure probability logic theories, replacing for some purposes the use of saturated models for first-order theories; this will be used in Section 5.

An elementary submodel M_0 is a kind of pool where everything not impossible has already happened. Finite measure, like compactness, constrains the breadth of possible phenomena from above, and together they lead to a well-understood theory over M_0 (higher de Finetti theorems, higher-dimensional Szemerédi lemma). We present a model-theoretic version in Appendix B, either over an elementary submodel (following Towsner) or assuming qualitative higher amalgamation of types. But we also pose the question of finding the essential structures governing higher independence and hidden within M_0. In this we try to emulate Shelah's definability theorems for stable theories; definability of types *over a model* is easier, but it was really the recognition of $\mathrm{acl}^{\mathrm{eq}}(0)$ and the proof of definability over that that enabled a useful theory of independence. We obtain a satisfactory result for $n = 3$, using auxiliary stable structures piecewise interpreted in the theory, so that an expectation statement can be referred to the stable structure, in this case Hilbert spaces.

These results will actually be required in the somewhat more general setting of local logic, where a large-scale metric is given and only balls of finite radius are assumed to have finite measure.

Appendix A develops basic stability theory for invariant relations, i.e., disjunctions of $\bigwedge$-definable relations; the specialization to $\bigwedge$-definable relations is used in Sections 2 and 3. Appendix C illustrates the use of probability logic in the setting of mixing results on groups over pseudofinite fields.

Many open problems are described throughout the text.

Related work. While I thought at first that this was new territory, I soon learned that approximate equivalence relations, by other names, are already very well studied. I talked about Theorem 5.5 in Aner Shalev's meeting on *Groups and Words*, in June 2012. Immediately afterwards, Nati Linial pointed out to me the relation of Theorem 4.3 with the work of Lovász and Szegedy [59; 60] on *graphons*. Indeed while the language is different and the assumptions are slightly different, I believe that basic methods of graphons yield an alternative proof of Theorem 4.3.

The "pure" probability logic we use, and the notion of ultraproduct that we obtain from it, are also closely related to Razborov's flag algebras [69].

Many variations on probability logic appear in the model-theoretic literature, implicitly and explicitly. The main results of [53] are formulated within infinitary logic $L_{\omega_1,\omega}$, whereas for us the use of compactness is essential. Our semantics is in fact identical to that of definable Keisler measures [45] (named after a different work of Keisler's). One of the variations reported on in [53], due to Hoover, is finitary, as well as the treatment in [33]; they adjoin $\{0, 1\}$-valued predicates

$$P_x^{>\alpha}\phi(x, y),$$

intended to indicate that the event defined by $\phi(x, b)$ has probability greater than or possibly equal to α. However in such a setting compactness would dictate the existence of an event with probability > 0 but $< 1/n$ for each n; this is an intermediate state between emptiness and probability zero. Such ghost predicates are difficult to control, and frequently lead to undecidability due to measure-zero phenomena that are not really intrinsic to the probabilistic viewpoint. The use of continuous logic is thus natural, being compatible with compactness and the standard interpretation of real-valued probability at the same time.

The papers [51; 64; 63; 49]. (in the setting of $\aleph_0$-categoricity) are closely related to our independence theorem for probability logic, Theorem 3.16. In particular, [51, Theorem 1.1 or 3.4] can be seen as special cases of Theorem 3.16(1); the measure-preserving action on X assumed there can be viewed as data for definability of a measure on a new stochastic sort X. In [65; 50; 16], the very interesting examples of 2-graphs and kay-graphs are analyzed, showing in particular that a naive generalization of measure independence to higher amalgamation cannot hold, and on the other hand (see [65, 7.2.1, 7.2.3]) that it does hold in certain circumstances, related to Theorem B.8.

Ibarlucía in [49] employs a method of using auxiliary piecewise-definable stable structure, developed independently but very similar to ours; see also [28].

In the asymptotically finite setting, a statement equivalent to the stabilizer theorem for groups was independently proved in [70], using a beautiful combinatorial argument. (See also [18].) It is not clear if this method applies to finite approximate equivalence relations too.

As far as I know, the nearest result to Theorem 5.5 on approximately homogeneous approximate equivalence relations is the paper [12] of Benjamini, Finucane and Tessera. Their main focus is on finite approximate equivalence relations that are *exactly* homogeneous for a group action; for these, they obtain results of the same strength as [18], in particular showing that the phenomenon exists essentially for nilpotent groups and their homogeneous spaces. But in the one-dimensional case, they also consider approximately homogenous relations, again with technically somewhat different definitions than we use here.

A recent remarkable work of Gowers and Long [34] offers a wider interpretation of "almost structure".

2. Preliminaries

2.1. ***Real valued continuous logic.*** Continuous model theory dates back to the book [21]; other roots lie in Robinson's nonstandard analysis of the same period, and their development (notably Henson's study of nonstandard hulls of Banach spaces) in the 1970s. Krivine studied a real-valued logic, and stability was introduced to the area in [57]. A modern version, with a full-fledged stability theory as well as simplicity and NIP, was introduced by Ben Yaacov and coworkers in [8; 5] and other articles; see especially [9].

In continuous logic, terms are defined in the same way as in first-order logic, but formulas ϕ are taken to take truth values in some compact Hausdorff space X_ϕ. Any continuous map on $c : X_\phi \times X_\psi$ to a compact space Y can be viewed as a connective, thus creating a new formula $c(\phi, \psi)$ taking values in Y. Any continuous map q from the (Hausdorff) space of nonempty closed subsets of X to a compact space Y induces a quantifier, taking formulas $\phi(x, y)$ with range X to formulas $(qx)\phi(x, y)$ with free variables y and range Y. The interpretation of $(qx)\phi(x, b)$ in a structure A is $q(\mathrm{cl}\{\phi(a, b) : a \in A\})$ (Chang and Keisler, 1966).

Specifically in continuous *real-valued* logic, the spaces X_ϕ will be closed intervals in $\mathbb{R}$, or occasionally in $\mathbb{R}\cup\{\infty\}$. The connectives can be restricted to $+$, $\cdot$, 1 (Stone–Weierstrass). The quantifiers can be restricted to min and sup, though it is not always best to follow this religiously. We view two languages as having the same expressive power if a formula of one can be uniformly approximated by a formula of the other, and vice versa. (We do not seek formula-to-formula equality.)

A *complete theory* is a specified value for each sentence (formula with no free variables). Similarly, if M is a structure, a *type* $p(x)$ *over* M is a specified value for each formula $\phi(x)$ with parameters from M.

Let A be a substructure of M. Formulas with parameters from A, and variable x, define functions on the set $S_x(A)$ of types $p(x)$ over A; we topologize $S_x(A)$ minimally so that they are all continuous, in other words as a closed subset of the product topology. Then $S_x(A)$ is compact, and any continuous function on $S_x(A)$ is uniformly approximated by formulas. If Z is a closed (respectively open) subset of $S_x(A)$, we call $\{m \in M : \mathrm{tp}(m/A) \in Z\}$ a $\bigwedge$-definable (respectively $\bigvee$-definable) subset of M; these notions should be used only in sufficiently saturated models, say ones where every type over A is realized.

If u is an ultrafilter on a set I and $(a_i : i \in I)$ is an I-indexed family of elements of a compact Hausdorff space X, there is always a unique $x \in X$ such that any neighborhood of x contains almost all a_i (according to u). This is denoted $\lim_{i \to u} a_i$.

An *ultraproduct* along u of structures A_i for a real-valued language L is defined in the usual way, except that the value of a (basic) relation is the *limit along* u of the values on the coordinates.

2.2. *Metrics.* There is often a distinguished binary formula ρ whose interpretation is a metric $\rho : A^2 \to \mathbb{R}$ and such that every term and every basic formula are uniformly continuous, by a prescribed modulus of continuity. In this case, one modifies the definition of the ultraproduct by identifying elements at distance zero. This is analogous to the situation with equality in 2-valued logic.

2.3. *Localities.* Note that the rules of continuous logic would force ρ to have bounded image; indeed for the discussion so far, there is no harm in replacing ρ by $\min(\rho, 1)$. Ben Yaacov [4] defines an unbounded version, where a fixed unary function (the *gauge*) controls locality; it is similar to many-sorted logic where quantifiers are restricted to finitely many sorts; but in place of a discrete set of sorts one has a continuous family. We will however be interested in homogeneous structures, with a single 1-type, and they are not compatible with unary functions. We will thus use a binary function ρ^*, satisfying the laws of a metric; it could be the same as the metric ρ, or distinct from it; in any case we are mostly interested in ρ near 0 (to determine a topology) and in ρ^* near ∞ (to determine a coarse structure or, for us, the notion of locality, i.e., a family of sets where model-theoretic compactness will hold).

We allow relations R to take unbounded values; but we assume that any basic relation comes not only with a modulus of continuity (with respect to ρ) but also with a bound $b(R)$ on the support of R, so that

$$R(x_1, \dots, x_n) \le b(\max_{i,j} \rho^*(x_i, x_j)).$$

Here b is a continuous function $\mathbb{R} \to \mathbb{R}$ with compact support. Similarly basic functions are assumed to take values at a bounded distance from their arguments. We redefine saturation by restricting to types at bounded distance.

When we take an ultraproduct, we have to make an additional choice, beyond that of the ultrafilter. Let $\mathcal{M}_0$ be the naive ultraproduct; then $\rho^*(x, y)$ can be infinite. The relation $\rho^*(x, y) < \infty$ is an equivalence relation. We make a choice of one class. Thus, in an ultraproduct, $\rho^*(x, y)$ is finite *by definition*. We refer to this as a local ultraproduct with locality relation ρ^*.

We note that this logical structure depends on ρ^* only up to coarse equivalence; replacing ρ^* by $j \circ \rho^*$, where $j : \mathbb{R}^+ \to \mathbb{R}^+$ is an order-preserving bijection, will make no difference.

See Section A.1 for more detail.

Example 2.4. The language of Hilbert spaces is taken to have sorts S_r for any real $r \ge 0$, denoting the ball of radius $\le r$. It has function symbols $0, +, \cdot\alpha$ for any $\alpha \in \mathbb{C}$, so $+ : S_r \times S_{r'} \to S_{r+r'}$ and $\cdot\alpha : S_r \to S_t$ whenever $r|\alpha| \le t$. There is one additional basic relation for the inner product, $(\ , \ ) : S_r \times S_{r'} \to [-rr', rr']$, and the obvious axioms. The metric is taken to be $|x - y|$, where $|x| = \sqrt{(x, x)}$.

The division into sorts adds somewhat artificial structure; a better approach is developed in [4]. For the actual use of Hilbert spaces in this paper this will not be essential, we can take either one.

Suppose however we wish to consider H as an affine space, without a distinguished 0. In this case it will not do to add sorts, whether discretely or continuously. Instead we use the locality function $\rho^*(x, y) = \|x - y\|$; in this case it happens to coincide with the metric. The effect is again to limit quantifiers to bounded balls, but the balls can be anywhere on H.

2.5. ***Cobounded equivalence relations and the logic topology.*** Let X be a $\bigwedge$-definable set. A $\bigwedge$-definable relation Λ is called *a cobounded equivalence relation* if in any model M, Λ defines an equivalence relation on $X(M)$, and $X(M)/\Lambda$ has cardinality bounded independently of M.

We have $\Lambda = \bigwedge_i \Lambda_i$ with Λ_i definable, such that all antichains of Λ_i are finite, of size bounded by some $\beta_i \in \mathbb{N}$. If $M \models T$ and $N \succ M$, we can pick in M a maximal antichain $c_1, \dots, c_{\beta_i}$ for Λ_i. If $a, b \in X(N)$ have the same type over M, or even just the same Λ_i-type over M for each i, then $(a, c_i) \in \Lambda_i$ iff $(b, c_i) \in \Lambda_i$. By maximality of the antichain, we do have $(a, c_i) \in \Lambda_i$ for at least one $i \le \beta_i$, and hence $(a, b) \in \Lambda_i \circ \Lambda_i$.

If $M \models T$ and $N \succ M$, and $a, b \in X(N)$ have the same type over M, then $a \Lambda b$. Hence we have a natural map $S_X(M) \to X/\Lambda$. The image set is the same as $X(N)/\Lambda$ for sufficiently saturated N, and we denote it simply by X/Λ. The surjective map $S_X(M) \to X/\Lambda$ induces a topology on X/Λ — the *logic topology* — which is compact and Hausdorff. It does not depend on the choice of M. Further, for any reduct M' of M such that Λ is still $\bigwedge$-definable in M', since the compact M-induced topology refines the M'-induced Hausdorff topology, they coincide; expansions do not change the space X/Λ.

For the same reason, we have a well-defined map $S_X^\Lambda(M) \to X/\Lambda$ (where S_X^Λ is the space of Λ-types, meaning Λ_i-types for each i), and it induces the same topology on X/Λ. Hence, if Y is a closed subset of X/Λ, then the pullback of Y to X is a $\bigwedge$-qf-definable set with parameters in M.

Let T be a complete theory, X a sort. There exists a unique *finest* cobounded $\bigwedge$-definable equivalence relation Λ of T. For this choice of Λ, the space X/Λ is called the space of compact Lascar types of T in the sort X, or the Kim–Pillay space $\mathrm{KP}_T(X)$ ([39]). Similarly for the qf KP-space.

The classes of Λ are called the compact Lascar types, or Kim–Pillay types.

Remark 2.6. Let E be any cobounded $\bigwedge$-definable equivalence relation; assume it is defined using formulas from a family Φ, e.g., $xEy \iff \phi(x, y) = 0$ for each $\phi \in \Phi$. Then X/E is a quotient of $\mathrm{KP}_T(X)$, and we can compare the quotient topology to the topology defined as above using the space of Φ-types alone. As the former is stronger and both are compact and Hausdorff, they must be equal. In other words, the image in X/E of any $\bigwedge$-definable subset of X is already the image of a Φ-$\bigwedge$-definable set.

2.7. *Stability.*

Definition 2.8 ([57]). A formula $\phi(x, y)$ is *stable* if for any model M and any elements a_i, b_i ($i \in \mathbb{N}$) of M, if $\lim_{i\to\infty} \lim_{j\to\infty} \phi(a_i, b_j)$ exists and equals α and $\lim_{j\to\infty} \lim_{i\to\infty} \phi(a_i, b_j) = \beta$, then $\alpha = \beta$.

The class of stable formulas $\phi(x, y)$ is easily seen to be closed under continuous connectives.

Lemma 2.9. *Let H be a Hilbert space,with elements $a_i, b_i (i \in \mathbb{N})$ of the unit disk. If* $\lim_{i\to\infty} \lim_{j\to\infty}(a_i, b_j)$ *exists then so does* $\lim_{j\to\infty} \lim_{i\to\infty}(a_i, b_j)$, *and they are equal.*

This lemma means that $(\ , \ )$ is a stable formula. Using quantifier elimination for Hilbert spaces, this easily implies that every formula is stable; but we will need this particular one. The significance of this was realized in [57] but also in [36]; see [6]. In the context of expectation quantifiers E_t that will soon be introduced, it implies that $E_t(f(x, t)g(x, t))$ is always a stable formula.

The following is a continuous logic version of Shelah's finite equivalence theorem (uniqueness of nonforking extensions over algebraically closed sets); see [8]. The continuity in each variable is the open mapping theorem, or the definability of types. Joint continuity does not hold in general.

Statement (1) below asserts that any value of $\phi(x, b)$ other than α causes forking, while (2) is a strong converse asserting that the value α can be taken simultaneously for any family of b's.

As usual we write p to denote the solution set of p, and $\alpha(a, b)$ for $\alpha(a/E, b/E')$.

Theorem 2.10 ([8]). *Let $\phi(x, y)$ be a stable formula on $P \times Q$. Then there exist cocompact $\bigwedge$-definable equivalence relations E on P and E' on Q, and a Borel function $\alpha : P/E \times Q/E' \to \mathbb{R}$, continuous in each variable and automorphism invariant, such that in any sufficiently saturated model M*:

(1) *In any prescribed E'-class there exists a sequence $(b_j : j \in \mathbb{N})$, such that for all $a \in P$, $\lim_{j\to\infty} \phi(a, b_j) - \alpha(a, b_j) = 0$. Equivalently, for any $\epsilon > 0$, for some $k \in \mathbb{N}$, for any $J \subset \mathbb{N}$ with $|J| \geq k$, and any $a \models p$, for some $j \in J$, $|\phi(a, b_j) - \alpha(a, b_j)| < \epsilon$.*

(2) *For any finite set $\{b_j : j \in J\} \subset Q$ and any $\epsilon > 0$, there exists $a \in P$ in any prescribed E-class with $|\phi(a, b_j) - \alpha(\mathrm{tp}(a), \mathrm{tp}(b_j))| < \epsilon$ for each j.*

(3) *Let $M_0 \prec M$; assume ϕ is quantifier-free. Then $\alpha(a, b)$ depends only on* $\mathrm{qftp}(a/M_0)$ *and* $\mathrm{qftp}(b/M_0)$.

Thus $\alpha(p, q)$ gives the generic or expected value for $\phi(a, b)$ when $a \models p$, $b \models q$, and any deviation from this value will cause dividing. A more general statement will be proved in Appendix A (see Theorem A.27).

For each complete type $p \subset P$, α is continuous as a function of two variables on $p \times Q$. But in general it is not continuous on $P \times Q$, even for the theory of pure equality augmented with infinitely many constants. To see how this may arise in a probabilistic setting, consider the random graph, with infinitely many distinguished constants, and with a measure giving independent probability $\frac{1}{2}$ to an edge. Then for two types $p(x)$, $q(y)$, and for $\phi(x, y)$ the probability that z is a neighbor of both x and y, we have $\alpha(p, q) = \frac{1}{4}$ *unless* $p \vdash x = c_n$ and $q \vdash y = c_n$ for the same n; this is not bicontinuous.

2.11. ***Topologies.*** We discuss here some elementary topology that will be needed later.

By a *pseudometric* on X we mean a function $d : X^2 \to \mathbb{R}$ with $d(x, y) = d(y, x) \geq 0$, $d(x, x) = 0$, and $d(x, z) \leq d(x, y) + d(y, z)$. There is a canonical map $j : X \to \overline{X}$ into a complete metric space, preserving d, with dense image; we refer to $\overline{X}$ as the *completion* of (X, d). Typically j is not injective.

Let $f : X^n \to \mathbb{R}$ be a function, uniformly continuous with respect to d; i.e., for all $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in X^n$ and $d(x_i, y_i) < \delta$. Then f induces a function $\bar{f} : \overline{X}^n \to \mathbb{R}$. Hence if X carries an L-structure for some continuous logic language L, an L-structure on $\overline{X}$ is canonically induced. All continuous logic formulas are preserved, i.e., $\phi(jx_1, \ldots, jx_n)^{\overline{X}} = \phi(x_1, \ldots, x_n)^X$. In particular, the axioms for expectation quantifiers (Section 3(1–4)) are preserved. Hence if X is a stochastic sort, then $\overline{X}$ becomes one too.

Assume X is a stochastic sort, with expectation quantifiers E. Write $\Lambda(x, y)$ if $j(x) = j(y)$; then Λ is a $\bigwedge$-definable. Assume further as in Section 2.5 that Λ is cobounded, so that $\overline{X}$ is compact. Then the expectation quantifiers on X induce a Borel measure on $\overline{X}$, equivalently a positive linear functional $\int$ on the Banach space of real-valued continuous functions on $\overline{X}$. Simply define $\int f = E(f \circ j)$. We use here the fact that $f \circ j$ is a (uniform limit of) parametrically definable functions; this in turn can be seen using Stone–Weierstrass, as the definable functions into $\mathbb{R}$ form an algebra and separate points on $\overline{X}$.

2.12. Let d be a pseudometric given by a formula of bounded real-valued continuous logic. Assume the equivalence relation E defined by $d(x, y) = 0$ is cobounded. Then d induces a metric on X/E, interpreted in any sufficiently saturated model. The metric is by definition continuous with respect to the logic topology on X/E. The latter being compact, it follows that d induces the logic topology; moreover, X/E is complete and hence coincides, as a metric space, with the completion $\overline{X}$.

This is valid locally in local continuous logic, with a metric ρ, when d is definable and hence subordinate to ρ. Namely, for any fixed $a \in X$, let B be a ball of some finite radius r around X. Then as above, the logic topology on B/E coincides with the topology induced by d. It follows that globally, the logic topology on X/E is locally compact, and induced by d. Returning to B, by the same argument, the metric topology on B/E induced by d also coincides with the logic topology obtained by considering only subsets of B defined with parameters in B. We will use this remark later on.

2.13. ***Isometry groups.*** Let Y be a locally compact metric space. The isometry group $G = \mathrm{Aut}(Y, d)$ is topologized by the compact-open topology, or uniform convergence on compacts. On Y, the topology agrees with the topology of pointwise convergence. This is because a compact C admits a finite set D_ϵ that is ϵ-dense in C; so if f, g are isometries and $d(f(x), g(x)) < \epsilon$ on for $x \in D_\epsilon$, then $d(f(x), g(x)) < 3\epsilon$ for all $x \in C$.

It is clear that left and right translations are continuous. Thus to check continuity of inversion and multiplication, it suffices to verify it at the identity element. If $g_i \to \mathrm{Id}_Y$, then for any a we have $g_i(a) \to a$, i.e., $d(g_i(a), a) \to 0$; since g_i is an isometry, $d(a, g_i^{-1}(a)) \to 0$ so inversion is continuous. If also $h_i \to 1$, then for any a we have $h_i(a) \to a$; by local compactness we may take all $h_i(a)$ in some compact neighborhood C of a; since g_i approaches Id uniformly on C, for any $\epsilon > 0$, for large enough i we have $d(g_i(y)), y) < \epsilon$ for all $y \in C$ (for large enough i); in particular $d(g_i h_i(a), h_i(a)) < \epsilon$, so $d(g_i h_i(a), a) < 2\epsilon$ for large i. It follows that G is a topological group.

The action $G \times Y \to Y$ is also easily seen to be continuous. Moreover, for $x_0 \in Y$, the map $G \to Y$, $g \mapsto gx_0$ is closed, since it suffices to check this after restricting to a closed bounded subset Y' of Y, and the preimage of Y' is compact. It follows that the stabilizer $G_{x_0} = \{g : gx_0 = x_0\}$ is a closed subgroup, Gx_0 is closed in Y, and G/G_{x_0} is homeomorphic to Gx_0

2.14. *Graphs and metrics.* A binary relation R on X, viewed as a graph, is *connected* if for any $x, y \in X$ there exist $n \geq 1$ and $x_1, \ldots, x_n \in X$ with $x = x_1, x_n = y$, and such that $R(x_i, x_{i+1})$ or $R(x_{i+1}, x_i)$ hold for $i < n$. In this case, for the least such n, we define $d_R(x, y) = n - 1$; d_R will be referred to as the metric associated to R.

In the lemma below two metrics appear, but all topological terms refer to (X, d).

Lemma 2.15. *Let (X, d) be a locally compact metric space. Let $\mathsf{R} \subset X^2$ be a closed binary relation, with (X, R) connected, and let d_{R} be the associated metric on X. Assume some d-ball is contained in a d_{R}-ball, and every d_{R}-ball has compact closure. Let $G = \mathrm{Aut}(X, d, \mathsf{R})$ be the group of isometries of (X, d) preserving R. Then G is a locally compact topological group. For any $x \in X$ and any compact $U \subset X$, $\{g \in G : g(x) \in U\}$ is compact.*

Proof. We saw above that G is a topological group, and that the topology given above agrees with pointwise convergence; we will use the latter description in order to reduce to compactness of product spaces.

To show that G is locally compact, it suffices to show that 1_G has a compact neighborhood. Let b_0 be an open d-ball around a_0 contained in a d_{R} ball. Let D_n be the d_{R}-ball around a_0 of radius n; then $b_0 \subset D_{n_0}$ for some n_0. Let $U_0 = \{g \in G : g(a_0) \in b_0\}$. Then U_0 is an open neighborhood of 1 in G. If $g \in U_0$, then $g(D_n) \subset D_{n+n_0}$; indeed $g(a_0) \in b_0$ so $d_{\mathsf{R}}(a_0, g(a_0)) \leq n_0$. If $d_{\mathsf{R}}(x, a_0) \leq n$, then $d_{\mathsf{R}}(g(x), g(a_0)) \leq n$ (it is here that we use the assumption that R, hence d_{R}, are preserved by G). So $d_{\mathsf{R}}(a_0, g(x)) \leq n + n_0$. Hence $U_0 \subset U_1$, where U_1 is the set of isometries of X satisfying $g(D_n) \subset D_{n+n_0}$ for every n. Now U_1 is compact since it embeds homeomorphically into a closed subset of the product space $\prod_n D_{n+n_0}^{D_n}$, mapping $g \to (g|D_n)_n$. And G is clearly a closed subgroup of the isometry group; so $G \cap U_1$ is compact too. $\square$

2.16. ***NIP.*** Let $R(x, y)$ be a formula, and let $S_x(A) = S_x^R(A)$ be the space of quantifier-free R-types in the variable x over a set A. For $R(x, y)$ taking values in $\{0, 1\}$, R has NIP (does not have the independence property) if $|S_x(A)|$ grows at most polynomially in $|A|$, i.e., $|S_x(A)| \le C \cdot |A|^k$ for some C, k and for all A. By a theorem of Sauer, Shelah, and Vapnik–Chervonenkis, this is equivalent to $|S_x(A)| < 2^{|A|}$ for $|A| > m$; the minimal such m is the Vapnik–Chervonenkis dimension, and we have $|S_x^R(A)| \le 2|A|^m$.

For an $\mathbb{R}$-valued formula $R(x, y)$, one says that R has NIP if for any fixed $\epsilon > 0$, $S_x^R(A)$ grows at most polynomially in $|A|$ up to ϵ-resolution; in other words $S_x^R(A)$ can be covered by polynomially many ϵ-balls, or equivalently admits at most polynomially many (in $|A|$) pairwise disjoint ϵ-balls. See [5]. This notion generalizes to general continuous logic (with values in compact spaces).

On the other hand, we will say that R has pNIP (of degree d) if for $|A|, |n| \ge d$, $S_x^R(A)$ can be covered by at most $(|A|n)^d$ $1/n$-balls. Thus the growth is polynomial not only in the base size but also in the resolution. pNIP relations are closed under connectives corresponding to Lipschitz functions $\mathbb{R}^k \to \mathbb{R}$.

Remark 2.17. If the relation $\phi(x; yz) = R(x, y) \le z$ has NIP, then R has pNIP. Indeed the $R(x; y)$ types, up to $1/m$-resolution, over a set $b_1, \dots, b_n$ can be viewed as ϕ-types over $b_1, \dots, b_n, 0, 1/m, \dots, 1$ so their number is polynomial in mn.

We will later need an effective version of the *uniform law of large numbers* of Vapnik–Chervonenkis. What is essential for us, to obtain finite packing dimension, is that N in Proposition 2.18 should grow at most polynomially with n. This already follows from [74, Theorem 2]. (In the notation there, set $\epsilon_0 = 1/(2n)$; one looks for a lower bound on l that ensures that the right hand side is < 1; this ensures not only existence but a nonzero percentage of N-tuples $c_1, \dots, c_N$ such that for each b there are at least $1/(2n)$ values of i with $R(c_i, b)$.) However we quote more precise bounds.

Proposition 2.18 ([38]). *Let (U, μ) be a probability space, and $R(u, b)$ be a relation with Vapnik–Chervonenkis dimension d (i.e., the family of events $\{u : R(u, b)\}$ has Vapnik–Chervonenkis dimension d). Then for any n there exist $c_1, \dots, c_N$, $N = 8dn\log(8dn) \le 24(dn)^2$, such that for any b with $\mu R(u, b) \ge 1/n$, for some i we have $R(c_i, b)$.*

Proof. This follows from Corollary 3.8 of [38], with $\delta = \frac{1}{2}$. In [38] the result is stated for μ a normalized counting measure on a large finite set, but as the bound does not mention the size of this set, the result immediately extends to all probability spaces by a standard approximation argument; see [48] around 2.7. □

Remark 2.19. Let μ be a measure on ϕ-types, not necessarily generically stable. It is shown in [45, Lemma 4.8(i)] that for any n and any model M there exist $c_1, \dots, c_N$

in an elementary extension M^* of M, such that for any b with $\mu R(u,b) \geq 1/n$, for some i we have $R(c_i, b)$. The proof yields a polynomial bound $N \leq O(n^\delta)$ for some δ.

2.20. ***Weakly random types.*** Let M be an $\aleph_1$-saturated model, and let μ be a finitely additive measure on formulas over M, or just on Boolean combinations of formulas $\phi(x,b)$. A ϕ-type p over M is called *weakly random* if any formula ψ in p has $\mu(\psi) > 0$. Let X be the compact Hausdorff space of weakly random global ϕ-types; μ induces a Borel probability measure on X. In any theory, a formula dividing (or forking) over $\varnothing$ has measure zero for any 0-definable measure. Thus a weakly random ϕ-type cannot fork over $\varnothing$. In particular if we fix a model M_0, it cannot fork over M_0. In a NIP theory, if a type p over a saturated model $M \succ M_0$ does not fork over M_0, then for any θ there exists a set of types over M_0 $I(\theta)$ such that $\theta(x,c) \in p$ iff $\mathrm{tp}(c/M_0) \in I(\theta)$. (See, e.g., [45, 2.11].) Hence the set of weakly random types has cardinality at most $\beth_2(|M_0| + |L|)$. In fact X is separable (see [25, 2.9 and 2.10]) but I am not sure if it is in general metrizable. It is so in the case of a smooth measure, or a generically stable type.

3. Probability logic

Many versions of probability logic were investigated by Keisler and Hoover (see [53]) following work of Carnap, Gaifman, Scott and Krauss; most of these were based on $L_{\omega_1,\omega}$. We will use here a first-order, real-valued version based on continuous logic. This enables the use of compactness, and in particular studying families of finite structures via their probability logic limits.

While it is possible to mix inf, sup quantifiers with probability quantifiers, we will be mostly interested in *pure probability logic*, where only probability quantifiers are used.

The flavor of this logic is determined by Hoover's quantifier-elimination (see Theorem 3.6) and, over a model, the independence theorem Theorem B.11. Roughly speaking the first result gives a quantifier-elimination to one block of quantifiers; the latter says that unexpected interaction among events can occur finitely many times, but is no longer unexpected once seen often enough. In the case of binary interactions, a much more precise description is available; see Theorem 3.16. What *can* be interpreted is a compact structure, the "core", with an action of a compact group on it. It can be viewed as the space of Lascar types of singletons. Each binary relation gives rise to a binary function on this core, making it into a compact structure, and the probability quantifiers induce a measure on it. Given this core, along with the natural map of the universe into it, all values of all formulas obtained using probability quantifiers are completely determined. See Corollary 3.21.

For binary languages, these results indicate that probability logic (at least over binary languages) has substantial descriptive value but limited interpretative strength. Interesting unary and "almost unary" relations can be interpreted using probability quantifiers (the almost unary ones are just unary if the Galois group of the theory is trivial). However no new binary or higher relations can be defined, beyond combinations of almost unary ones and the originally given quantifier-free formulas. This is a severe restriction on the interpretative power of pure probability logic. It stands in contrast to the limitless interpretative abilities of first-order logic.

We will extend this to local probability logic, involving a locally compact core and a locally compact group acting on it.

By a *probability quantifier in x* we mean a syntactical operation from formulas $\phi(x, y)$ (with y a sequence of variables distinct from x) to formulas $E_x\phi$ in the variables y[3] satisfying:

(1) $E_x(1) = 1$ for the constant function 1 (viewed as a function of any set of variables).

(2) $E_x(\phi+\phi') = E_x(\phi)+E_x(\phi')$, and $E_x(\psi\cdot\phi) = \psi\cdot E_x(\phi)$ when x is not free in ψ.

(3) $E_x(|\phi|) \geq 0$.

Note that (1–3) are universal, first-order axioms.

If quantifiers are used, (3) (applied to $(\sup_x \phi - \phi)$) becomes equivalent to

(4) $E_x(\phi) \leq \sup_x \phi$.

However we will be interested especially in formulas of pure probability logic ϕ that do not involve quantifiers, so that it is preferable to have (4) explicitly.

Remark 3.1. If axioms (1)–(3) hold for pure probability logic formulas, (4) will hold in existentially closed models for this sublanguage. To see this let L_{pr} denote the formulas obtained from basic ones using expectation quantifiers alone; view all formulas of L_{pr} as basic. Let M be an existentially closed model for an L_{pr}-universal theory T_{pr} including axioms (1–3). If (4) fails, then for some $\epsilon > 0$, with $\rho(y) = E_x(\phi(x, y)) - \epsilon$, $\phi(x, y) \leq \rho(y)$ must follow from some $\theta(y, y')$, where $\theta(b, b')$ holds for some b, b' from M. But then $E_x\phi(x, y) \leq E_x\rho(y) = \rho(y) = E_x(\phi(x, y)) - \epsilon$, a contradiction. Thus (1)–(3) suffice to axiomatize the pure probability logic validities, i.e., the universal sentences applied to formulas using E_x but no other x-quantifiers.

The name *probability quantifier* arises from the situation where ϕ takes values in $\{0, 1\}$; in general one might also call it an *expectation quantifier*.

[3]We assume at the syntactical level that if x' is another variable of the same sort as x, and ϕ' is obtained from ϕ by using x' in place of x, then $E_x\phi = E_{x'}\phi'$.

3.2. ***Semantics.*** If M is any model of axioms (1)–(4), and $S_x(M)$ is the type space over M in variable x, we obtain a positive linear functional on a dense subset of $C(S_x(M))$, namely the interpretations of formulas $\phi(x)$ with free variable x and parameters in M; it follows from the last axiom that if $\phi(x, b)$ defines the same function as $\phi'(x, b')$, then $E_x\phi(x, b) = E_x\phi'(x, b')$.

By the Riesz representation theorem, there exists a unique regular Borel measure μ_x on $S_x(M)$ with $\int \phi(x)\, d\mu(x) = (E_x\phi)$. *This is the intended semantics, and constitutes the completeness theorem for expectation logic.*

For pure probability logic, we take $S_x(M)$ to be the quantifier-free type space.

For example, the volume of an r-ball B around b is determined by the pure probability logic type of b. This is evident from the semantics as the Borel measure on the type space over M includes this information; B is a $\bigwedge$-definable set, and corresponds to a closed subset of $S_x(M)$. One can also see this directly, but more computationally, by expressing vol(B) as the limit of $E_x\theta(d(x, b))$, where θ is a continuous function into [0, 1] supported on $[0, r]$, approximating the characteristic function of $[0, r]$ in the uniform norm.

By a *stochastic sort*, we mean a sort endowed with such an operation $\phi \mapsto E_x\phi$. We will not necessarily assume that every sort is stochastic.

3.3. ***Fubini.*** If X and Y are stochastic sorts, with corresponding probability quantifiers E_x and E_y, we obtain two measures on the variables (x, y), arising from E_xE_y and E_yE_x. They agree on formulas obtained by connectives from formulas in x and formulas in y. On *compact* sorts, this suffices to force the two measures to commute. In general they may not, even if $X = Y$, since a function defined by ϕ on $S_{x,y}$ may not be measurable for the product measure. We will say that E_x, E_y commute if $E_xE_y\phi = E_yE_x\phi$ for all ϕ; see [48]. We say that Fubini holds if any two stochastic sorts commute.

On the other hand, in the foundations of NIP theories notably, one encounters Keisler measures that do not commute. Thus we do not include Fubini in the list of axioms, but invoke the assumption when needed.

3.4. ***Pseudofinite semantics.*** The above treatment of probability logic takes as a starting point a family of formulas, closed under an expectation operator, as well as continuous connectives. This is analogous to a view of logic as a family of formulas, closed under quantifiers and connectives. In another approach, one forms the family of formulas formally by closing the basic relations under continuous connectives, the $\inf_x$ operator and the (Ez) operator for stochastic sorts. Each basic formula R comes with a real interval I_R (so that R takes values in I_R) and a uniform continuity modulus μ_R, so that $|R(x_1, \ldots, x_n) - R(y_1, \ldots, y_n)| \leq \mu_R(\max d(x_i, y_i))$. These are propagated to general formulas ϕ in the natural way; in particular the interval and uniform continuity modulus of $(Ez)\phi$ and of $\sup_z \phi$ are defined to be those of ϕ. Let us explain, given a finite structure or an ultraproduct M of finite structures, how to evaluate each formula ϕ. This is done by induction on the complexity of formulas. The value $(Ez)\phi(z, m)$ is defined to be the mean value of $\phi(z, m)$ (with respect to the counting measure on the relevant sort of M). The values of $\inf_x \phi$ and $C(\phi_1, \ldots, \phi_k)$ (where $C : \mathbb{R}^k \to \mathbb{R}$ is continuous) are also defined in the obvious way.

Note that in a saturated model, where every type avoiding measure-zero formulas is realized, it may be impossible to avoid nonempty parametrically definable measure zero sets. In this case, the pure probability logic type of an element in this theory need not determine the isomorphism type.

3.5. ***Hoover's normal form.*** We return to the general setting of probability logic. Here is Hoover's theorem on reducing expectation quantifiers to a single block; it is valid in general in our setting with several stochastic sorts. See [53] (in a slightly different setting).

Theorem 3.6 (Hoover). *Any formula $\psi(y)$ built using connectives and expectation quantifiers can be approximated by ones of the form $E_x\phi$, where $\phi(x, y)$ has no* (*probability*) *quantifiers, and x is a sequence of variables.*

Proof. Let Ψ be the class of formulas that can be so approximated. Clearly Ψ contains the quantifier-free formulas, and is closed under probability quantifiers; we have to show in addition that Ψ is closed under connectives corresponding to continuous functions c. We give two proofs of this.

The first works directly for any c. Let $\bar{y}$ be a sequence of a large number N of copies of y. By the law of large numbers, $E_y\phi$ is approximated by $\frac{1}{N}\sum_j \phi(y_j)$, uniformly in the remaining free variables of ϕ. (ϕ takes values in a bounded interval, say $[0, 1]$; so $|\phi - E\phi| \leq 1$ and thus $(\phi - E\phi)^2$ has expectation at most 1. A weak version of the law of large numbers now states that $\left|\frac{1}{N}\sum_j \phi(y_j) - E(\phi)\right| \leq \lambda$ with probability at least $1 - 1/(N\lambda^2)$. Taking $\lambda = N^{-1/4}$ will do.) So $c(E_y\phi)$ is approximated by $E_{\bar{y}}c\left(\frac{1}{N}\sum_j \phi(y_j)\right)$, uniformly in the other variables.

The second proof was explained to me by Itaï Ben Yaacov. It does not require Fubini. Using the Stone–Weierstrass theorem we may take c to be a polynomial. This decreed, no further approximations are needed; the normal form becomes valid purely algebraically. We have to consider the sum or product of two expressions $E_x\phi$, $E_y\psi$ where we may assume the quantified variables x, y are disjoint from each other and the free variables. In this case the sum is $E_xE_y(\phi + \psi)$ and the product is $E_xE_y\phi \cdot \psi$. □

3.7. ***Stability of binary correlations.*** Ben Yaacov proved the stability of the theory of measure algebras in [3]. Taking the viewpoint of piecewise interpretable structures — in this case measure algebras — this immediately implies the stability of $E_x\phi(x, y) \wedge \psi(x, z)$ in any theory with a real-valued expectation operator. The implication was not immediately noticed, however, and stability of $E_x\phi(x, y) \wedge \psi(x, z) = 0$ was reproved directly in [24] in a restricted environment, in order to prove the independence theorem there. This was then transposed to the forking ideal in place of the measure 0 ideal in [40] (in finite S1-rank) and [54] for general simple theories, yielding the independence theorem for simple theories. Here we return to measure correlation and give the simple proof from [3].

Proposition 3.8. *For any $\phi(x, y)$ and $\psi(x, z)$ valued in $\{0, 1\}$, the formula $\theta(y, z)$ defined by $E_x\phi(x, y) \wedge \psi(x, z)$ is stable.*

More generally, for any $\phi(x, y)$ and $\psi(x, z)$, the formula $E_x(\phi(x, y)\cdot\psi(x, z))$ is stable. In fact, for any formulas $\phi(x, y)$ and $\psi(x, z)$ valued in a compact subset C of $\mathbb{R}$, and any continuous function $c : C^2 \to \mathbb{R}$, the formula $E_x(c(\phi(x, y), \psi(x, z)))$ is stable.

Proof. We prove the second statement first. Let M be a model, and let $b_i, c_j \in M$. Let S be the type space over M in the variable x. The expectation quantifiers induce a measure μ on S such that $\int \phi(x)\, d\mu(x) = (E_x\phi)$ for any formula $\phi(x)$ over M. Now $\phi(x, b_i)$ defines a continuous, bounded real-valued function f_i on S, while $\psi(x, c_j)$ defines g_j. So $f_i, g_j \in L^2(X, \mu)$, and $E_x(\phi(x, b_i) \cdot \psi(x, c_j)) = \int f_i g_j = (f_i, g_j)$. Thus stability follows from Lemma 2.9.

The first statement is a special case, since $\wedge = \cdot$ on $\{0, 1\}$.

As for the third statement, we can approximate c uniformly by a polynomial; so we may take c to be a polynomial. Since E_x is additive, we may take c to be a monomial $p_m(u)p_n(v)$, where p_n denotes the n-th power map. Replacing ϕ by $p_m \circ \phi$ and ψ by $p_n \circ \psi$, the statement now follows again from the first paragraph. $\square$

Let Y be the sort of the variables y, and let $\hat{Y}$ denote the associated strong type spaces, i.e., $\hat{Y} = Y/E$ with E the smallest $\bigwedge$-definable cocompact equivalence relation. Let $\hat{b}$ denote the image of b in $\hat{Y}$; similar notation for z, Z.

By Theorem 2.10, it follows that there exists a function $\alpha : \hat{Y} \times \hat{Z} \to \mathbb{R}$, continuous in each variable, such that for any $b \in Y$, $c \in Z$ with $c \perp z$ we have

$$E_x(\phi(x, b) \wedge \psi(x, c)) = \alpha(\hat{b}, \hat{c}).$$

3.9. ***The independence theorem and statistical independence.*** Proposition 3.8 and Theorem 2.10 combine to yield a basic principle of probability logic, *the independence theorem*. A qualitative version is true in greater generality for ideals with a certain saturation property, (S1), enjoyed by the measure-zero ideals of measures. As noted in [43], when we actually have a measure it is possible not only to assert that the value of $\mu(R(a, z) \cap R'(b, z))$ is uniquely determined, but to give an explicit formula for it. We include a proof here, though the independence theorem will only be used in qualitative form later on.

Towsner [73] noted the relation to combinatorial results and gave a proof of a related statement of "triangle-removal" type for n-amalgamation over a model, by L^2-methods.

The proof given below reconciles these two approaches. The structure interprets (piecewise) a Hilbert space, where stability reigns in the qf part; this can be viewed as the true source of the stability of the formula in question. The parametrizing sorts are not required to carry a measure, and the exceptional set is recognized explicitly.

We assume Z is a sort with expectation operators, X, X' are two other sorts, $R \subset X \times Z$ and $R' \subset X' \times Z$ are two relations. For some purposes we will assume that X, X' also carry expectation operators, and that Fubini holds; this will be stated explicitly.

Let us say that two functions f, f' on a measure space (X, μ) are independent if for any Borel B, $B' \subset \mathbb{R}$, $f^{-1}(B)$ and $(f')^{-1}(B')$ are statistically independent events. Equivalently, for any two bounded Borel functions e, e' on $\mathbb{R}$, $E((e \circ f) \cdot (e' \circ f')) = E(e \circ f)E(e' \circ f')$. If f and f' are characteristic functions of two events, this is the usual notion of statistical independence.

We say that f, f' are independent *over* a σ-subalgebra $\mathcal{B}$ of the measure algebra if for each such e, e', the conditional probabilities relative to $\mathcal{B}$ satisfy

$$E((e \circ f) \cdot (e' \circ f') : \mathcal{B}) = E(e \circ f : \mathcal{B}) E(e' \circ f' : \mathcal{B}).$$

A suggestive case occurs topologically when $\pi : X \to Y$ is a continuous map of Polish spaces, μ_y is a Borel family of measures on the fibers, $\nu = \pi_* \mu$, $\mathcal{B}$ is the measure algebra of ν, and $\mu = \int_y \mu_y$; this means that for any continuous function ϕ on X we have $\int \phi \, d\mu(x) = \int \big(\int \phi(x) \, d\nu_y(x) \big) \, d\nu(y)$. In this case f, f' are independent iff for almost all $y \in Y$, f, f' are independent with respect to μ_y. We will also say in this case that f, f' are statistically independent over Y.

Definition 3.10. Write $A \mathop{\smile\!\!\!\!\!|}_C B$ if for any stable continuous logic formula $\phi(x, y)$ over C, and tuples a from $A \cup C$, b from $B \cup C$, $\mathrm{tp}(a/b)$ does not ϕ-divide over C. In other words, if $\phi(a, b) = \alpha$, then for any indiscernible sequence $(b, b_1, b_2, \dots)$ over C and any $\epsilon > 0$ and n, there exists a' with $|\phi(a', b_i) - \alpha| < \epsilon$ for $i \leq n$. If $C = \varnothing$, we write $A \mathop{\smile\!\!\!\!\!|} B$.

If we restrict to stable formulas $\phi(x, y)$ with ϕ defined over $\varnothing$, write $A \mathop{\smile\!\!\!\!\!|}_{0;C} B$.

We use continuous logic formulas in this definition even if T is a first-order theory.

Let $R(x, y) = \bigwedge_i R_i(x, y)$, where the family $\{R_i : i \in I\}$ can be taken to be closed under finite conjunctions. Assume $R(x, b)$ divides, i.e., there exists an indiscernible sequence $(b, b_1, b_2, \dots)$ such that $\bigcap_j R(x, b_j)$ is inconsistent. Let μ be a definable measure, or more generally an invariant measure (i.e., $\mu(\phi(x, c))$ depends only on

ϕ and on tp(c)). Then $\mu(R(x,b)) = 0$, in the strong sense that $\mu(R_i(x,b)) = 0$ for some $i \in I$. See [43, 2.9]. Thus if $\psi(x)$ is a definable set of positive measure, then for any B, in some elementary extension, there exists a with $\psi(a)$ and such that no $R(a,b)$ holds, with $b \in B$, if R is a $\bigwedge$-definable relation such that $R(x,b)$ divides. This can be applied to any specific value α (or any closed range of values) of a continuous logic formula $\phi(x,y)$, letting $R(a,b)$ hold if $\phi(a,b) = \alpha$.

Let $\mathrm{cl}(A)$ denote the bounded closure, or continuous-logic algebraic closure in the sense of [8]. Thus a type over $\mathrm{cl}(A)$ is the same as a Kim–Pillay strong type over A.

Lemma 3.11. *Stable independence has the properties of*

- *symmetry*: *$A \mathop{\smile\!\!\!\!\!\!|}_C B$ implies $B \mathop{\smile\!\!\!\!\!\!|}_C A$.*
- *trivial monotonicity*: *$A \mathop{\smile\!\!\!\!\!\!|}_C B$ implies $A \mathop{\smile\!\!\!\!\!\!|}_C B'$ if $C \subseteq B' \subseteq B$. Also, $A \mathop{\smile\!\!\!\!\!\!|}_{0;C} B$ implies $A \mathop{\smile\!\!\!\!\!\!|}_{0;C'} B$ if $C \subseteq C' \subseteq B$.*
- *finite character.*
- *small bases*: *For any A, B there exists $C \subseteq B$, $|C| \leq |A| + |L|$ with $A \mathop{\smile\!\!\!\!\!\!|}_C B$.*
- *existence and stable stationarity*: *For any b and any Kim–Pillay type Q (over C), there exists $a \in Q$ with a, b stably independent over C. Moreover for any stable formula $\phi(x,y)$ over C, the truth value of $\phi(a,b)$ is the same for all such a.*
- *transitivity*: *If a, b are stably independent over A and a, c are stably independent over* $\mathrm{cl}(A \cup \{b\})$, *then a is stably independent over A from (b, c).*

Proof. All but transitivity follows directly from [8]. (Transitivity is proved there under the assumption of global stability; we check it here under our more local assumptions.)

So, let us prove transitivity. Work over A. Let $\phi(x; y, z)$ be stable. Note that any instance $\phi(x; b, z)$ is stable (there are no sequences (a_i, b_i, c_i) with $\phi(a_i; b_j, c_j)$ iff $i < j$; in particular no such sequences with all $b_j = b$). Let $p = \mathrm{tp}_{\mathrm{KP}}(a)$, and let $\psi(y,z)$ be the p-definition of ϕ. Then $\psi_b = \psi(b,z)$ is the $\mathrm{tp}(a/\mathrm{cl}(b))$-definition of $\phi_b(x,z) = \phi(x; b, z)$. This is because there exists a ("Morley") sequence a_i such that for any (b', c'), $\lim_i \phi(a_i, b', c') = \psi(b', c')$; in particular this holds for $b' = b$. On the other hand the existence of such an indiscernible sequence implies that for any $\beta \neq \psi(b, c')$, $\phi(a, b, y) = \beta$ divides for $y \models \mathrm{tp}(c'/b)$; so the $\mathrm{tp}(a/\mathrm{cl}(b))$-definition of $\phi_b(x,z)$ must be $\psi(z)$.

Now assume a, b are stably independent over A and a, c are independent over $\mathrm{cl}(Ab)$. Then $\phi(a; b, c)$ holds iff $\psi_b(c)$ iff $\psi(b, c)$. This shows the stable independence of a and (b, c). $\square$

3.12. ***Analytic structures viewed as interpretable.*** Let μ be a definable measure (in variable x) for a theory T, for instance obtained using expectation quantifiers. We will view the Hilbert space $L^2(\mu)$ as piecewise-interpretable in T. For any model $M \models T$, we have $L^2(\mu)(M) = L^2(S_x(M), \mu)$.

Then the Hilbert space formulas provide us with stable formulas of T, in the sense of continuous logic; and the results on stable independence apply.

For our purposes, we could use the theory of probability algebras in place of the theory of Hilbert spaces; the probability algebra $B(\mu)$ can be identified with the elements of $L^2(\mu)$ represented by $\{0, 1\}$-valued Borel functions, but we have not only the induced norm from $L^2(\mu)$ but also *multiplication* as part of the structure. Stability of $B(\mu)$ can in any case be deduced from that of $L^2(\mu)$. For other applications, the Banach lattice $L^1(\mu)$ will be needed. The Hilbert space picture is appealing in particular in connection with the Peter–Weyl theorem, and the representation of the automorphism group of the algebraic (bounded) hyperimaginaries. For definiteness we will talk about Hilbert spaces below, but the discussion would be the same for the others.

Let $\mathbb{U}$ be a large saturated model of T. Let H be the Hilbert space $H = L^2(S_X(\mathbb{U}))$. For any small substructure A, possibly including (hyper)imaginaries, we define H_A to be the subspace of H fixed by $\mathrm{Aut}(\mathbb{U}/A)$.

Remark 3.13. There is a canonical embedding of $L^2(S_X A)$ into H, falling into H_A (namely, $f \mapsto f \circ r$, where r is the restriction $S_X(\mathbb{U}) \to S_X(A)$). This will be surjective assuming a certain "strong germ" property; without such an assumption, there may exist a family $(D_c : c \in Q)$ of definable sets such that $\mu(D_c \triangle D_{c'}) = 0$ for all $c, c' \in Q$, but no A-definable set is equivalent to any D_c. We will not make this assumption, so the image of $L^2(S_X A)$ in H may be smaller than H_A.

The elements of H can be viewed as hyperimaginaries of $\mathbb{U}$; in fact H is piecewise interpretable in $\mathbb{U}$, in a sense that we now explain. An element ξ of $L^2(S_X(\mathbb{U}))$ can be approximated by continuous functions, given as the value of a formula $\phi(x, a)$ of real-valued continuous logic. Thus

$$\xi = \lim_n \phi_n(x, a_n)$$

with the limit taken in the L^2-norm. Let $\kappa = \|\xi\|_2$ and $\phi_n(x, a_n)$ of L^2-norm $\leq k+1$. Moreover we can choose the sequence with $\|\phi_n(x, a_n) - \xi\|_2 \leq 2^{-(n+1)}$, so that

$$\|\phi_n(x, a_n) - \phi_n(x, a_{n+1})\|_2 \leq 2^{-n}, \qquad \|\phi_n(x, a_n)\|_2 \leq \kappa + 1.$$

Let $\bar{a}$ be the sequence (a_n), and $\bar{\phi}$ the sequence ϕ_n, and write $L_{\bar{\phi}}$ for the set of sequences $\bar{a}$ satisfying the displayed formula. The number of possibilities for $\bar{\phi}$ is bounded (by $|L|^{\aleph_0}$); the sets $L_{\bar{\phi}}$ are easily seen to be directed, under inclusion (by using disjunctions, and using parameters to choose the appropriate disjunct). For each such $\bar{\phi}$, $L_{\bar{\phi}}$ is a $\bigwedge$-definable set. In this sense, $\bigcup_{\bar{\phi}} L_{\bar{\phi}}$ is piecewise $\bigwedge$-definable. The equivalence relation $\lim \phi_n(x, a_n) = \lim \phi_n(x, b_n)$ is also $\bigwedge$-definable; it is equivalent to

$$\bigwedge_n E_x\big((\phi_n(x, a_n) - \phi_n(x, b_n))^2\big) \leq 2^{-(n-2)},$$

where E_x is the expectation. Moreover, within $L_{\bar{\phi}}$, the relations $a + b = c$, $\alpha a = a'$ (for $\alpha \in \mathbb{R}$) are $\bigwedge$-definable, and so is the formula giving the inner product; it is approximated by $E_x \phi_n(x, a_n)\phi_n(x, b_n)$, uniformly in (a, b) for a, b such that $\|a\|_2$, $\|b\|_2$ is bounded.

Let L_H be the language of H, i.e., in our case the Hilbert space language.

Lemma 3.14. *Assume $A \leq B \cap C$, $A = \mathrm{cl}(A)$ and B, C are stably independent over A. Let H be a stable structure, piecewise interpretable in the sense considered above. Assume H_A is closed in H, i.e., every H-hyperimaginary element that is bounded over H_A lies in H_A. Then H_B and H_C are L_H-independent over $H'_A := H^{\mathrm{eq}} \cap A$.*

Proof. Let $b \in H_B$, $c \in H_C$ and let $\phi(x, y)$ be a (stable) L_H-formula. We view b, c as hyperimaginary elements of M. Then $\mathrm{tp}(b/A)$ is consistent with an A-definable ϕ-type $p(x)$. Note that the canonical base of p, in the sense of L_H, is then defined over A and lies in H^{eq}, so p is defined over H'_A. By the stable independence of B, C over A, $\phi(b, c)$ holds iff $\phi(x, c) \in p|C$. As this is true for every L_H-formula ϕ, b, c are L_H-independent over H'_A. □

In the case of Hilbert spaces, it is known [7] that for any imaginary h, we have $\mathrm{cl}(h) = \mathrm{cl}(e)$ with e coding a finite-dimensional subspace E of H. (Note that the lemma refers to hyperimaginaries of H itself, not to the induced structure from the piecewise-interpretation.) Moreover, $E \subseteq \mathrm{cl}(h)$. (It suffices to show that $E_1 := \{a \in E : \|a\|_2 = 1\} \subset \mathrm{cl}(e)$. But E_1 is compact, and e-definable, so by definition of "closed" for continuous logic, $E_1 \subseteq \mathrm{cl}(e)$.) Thus we obtain a simpler form, Lemma 3.15.

Lemma 3.15. *Assume $A \leq B \cap C$, $A = \mathrm{cl}(A)$ and B, C are stably independent over A. Then H_B and H_C are independent as Hilbert spaces over H_A.*

Let Z be a stochastic sort. We have a regular Borel measure μ on $S_Z(\mathbb{U})$ induced by the Z-expectation operators. Let H be the associated piecewise-hyperdefinable Hilbert space, and B the associated probability algebra. Let $H_{\tilde{0}}$, $B_{\tilde{0}}$ be the subspace (respectively subalgebra) of H, B consisting of elements with bounded orbit over $\varnothing$ (almost invariant). We will also write simply $\tilde{0}$ for $B_{\tilde{0}}$; so $H_{\tilde{0}}$ can be identified with $L^2(B_{\tilde{0}})$.

$B_{\tilde{0}}$ consists of a bounded number of elements, each of which lives in some hyperdefinable piece of B and hence can be identified with a hyperdefinable element.

Hence two elements with the same Kim–Pillay type have the same type over $B_{\tilde{0}}$; the equivalence relation $\mathrm{tp}(x/B_{\tilde{0}}) = \mathrm{tp}(y/B_{\tilde{0}})$ is $\bigwedge$-definable, and with boundedly many classes.

Let $\hat{X}$ denote the set of Kim–Pillay strong types of elements of X; so $\hat{X} = X/E$ for a certain cobounded $\bigwedge$-definable equivalence relation on X. For $a \in X(\mathbb{U})$, let $\hat{a}$ denote the image of a in $\hat{X}$.

Since $B_{\tilde{0}}$ consists of boundedly many hyperimaginaries, $\hat{a}$ determines $\mathrm{tp}(a/B_{\tilde{0}})$.

Let $\Xi = \hat{Z}$ be the maximal bounded quotient of Z by a $\bigwedge$-definable equivalence relation. Ξ carries a Borel probability measure, the pushforward μ_0 of μ on $S_Z(\mathbb{U})$ (or any $S_Z(M)$).

Recall that we write $a \mathop{\smile\!\!\!\!\!\!\,|}\, b$ if $\mathrm{tp}(a/b)$ does not divide over $\varnothing$ with respect to any stable formula.

Theorem 3.16 (independence theorem for probability logic). *Let (Z, μ) be a stochastic sort.*

(1) *Let $a \in X(M)$, $b \in Y(M)$, with $a \mathop{\smile\!\!\!\!\!\!\,|}\, b$. Then $R(a, z)$, $R'(b, z)$ are statistically independent over $\tilde{0}$.*

(2) *Assume X carries a definable measure ν, and ν, μ commute (i.e., Fubini holds). Then there exist a $\bigwedge$-definable set $X^* \subset X$ of full measure, such that if $a \in X^*$, $b \in Y(M)$, and $a \mathop{\smile\!\!\!\!\!\!\,|}\, b$, then $R(a, z)$, $R'(b, z)$ are statistically independent over Ξ.*

In the proof below, we will use the measure space $(X, \tilde{0}, \mu|\tilde{0})$; we write an integral of a measurable function f with respect to this space simply as $\int_{\tilde{0}} f$. We will also use the compact space $\hat{X}$; as a measure space, it is always understood to carry the measure derived by pushforward from ν (and also denoted ν). Ξ always carries the pushforward measure from $S_Z(\mathbb{U})$. We denote conditional expectation with respect to a subalgebra A by $E(f|A)$; when A is the measure algebra of a space S, e.g., of $S = \hat{X} \times \Xi$, we will also write $E(f|S)$ for $E(f|A)$.

Proof of Theorem 3.16(1). For $a \in X(M)$, define a function $\hat{a} : S \to \mathbb{R}$, $q \mapsto R(a, c)$, where $c \models q$; similarly $\hat{b}$. Let e, e' be bounded Borel functions on $\mathbb{R}$, and write a^e for $e \circ \hat{a}$. Let $E_r(\phi) = E(\phi|\tilde{0})$ denote the conditional expectation to $\tilde{0}$. We have to show that as $\tilde{0}$-measurable functions, we have $E_r(a^e \cdot b^{e'}) = E_r(a^e)E_r(b^{e'})$. Equivalently, for any bounded $\phi \in B_{\tilde{0}}$, integrating with respect to $(\tilde{0}, \mu|\tilde{0})$ we have

$$\int \phi(r) E_r(a^e \cdot b^{e'}) = \int \phi(r) E_r a^e E_r b^{e'}.$$

We can restrict to the characteristic functions of clopen sets, and view ϕ as a $\{0, 1\}$-valued, almost invariant Borel function $S_z(M)$.

Let a_0 be the orthogonal projection to $H_{\tilde{0}}$ of ϕa^e, $b_{\tilde{0}}$ the orthogonal projection to $H_{\tilde{0}}$ of $b^{e'}$. For any element $g \in H_{\tilde{0}}$, we have $\int g(r)\phi(r)E_r a^e = \int g \circ \pi(r)(\phi \circ \pi) a^e =$

$(g, \phi a^e)_H = (g, a_0)_{H_{\tilde{0}}}$. So $a_0(r) = E_r(\phi(r)a^e)$. Similarly, with the natural notation, $b_0(r) = E_r(b^{e'})$. Thus what we have to show is

$$\int \phi(r)a^e \cdot b^{e'} = \int a_0(r)b_0(r),$$

or in Hilbert space notation,

$$(\phi a^e, b^{e'})_E = (a_0, b_0)_{H_{\tilde{0}}}.$$

Now $\phi a^e - a_0 \perp H_{\tilde{0}}$, and $b^{e'} - b_0 \perp H_{\tilde{0}}$. In particular,

$$(\phi a^e - a_0, b_0) = (a_0, b^{e'} - b_0) = 0.$$

Thus what we need is that $\phi a^e - a_0$ and $b^{e'} - b_0$ are orthogonal vectors. This follows from Hilbert space independence of $a^e, b^{e'}$, which we have by Lemma 3.15. □

The proof of Theorem 3.16(2) involves three steps:

(a) Define a subset X^* of X, consisting of the elements whose R-interaction with $\tilde{0}$ factors through Ξ, and show it is $\bigwedge$-definable.

(b) Show $\nu(X^*) = 1$. It will be useful to prove this under a weaker assumption than stated, namely that ν is $\hat{X}$-definable rather than 0-definable.

(c) For $a \in X^*$ and any $b \in Y$ with $a \downarrow b$, $R(a, z)$, $R(b, z)$ are statistically independent over Ξ.

Towards (a), let β_a be the orthogonal projection of $R(a, z)$ to $H_{\tilde{0}}$.

We noted earlier, using the fact that the elements of $B_{\tilde{0}}$ are hyperdefinable, that $B_{\tilde{0}} \subset \mathrm{dcl}(\Xi)$. Hence β_a depends only on $q = \mathrm{tp}(a/\Xi)$, and thus only on $\hat{a}$. We may also write $\beta_{\hat{q}}$.

Let $\alpha_{\hat{a}}$ be the orthogonal projection of $R(a, z)$ (or equivalently of β_a) to $L^2(\Xi)$. Define

$$X^* := \{a \in X : \beta_a \in L^2(\Xi)\}.$$

Equivalently, $a \in X^*$ if $\alpha_{\hat{a}} = \beta_{\hat{a}}$.

Lemma 3.17. *X^* is $\bigwedge$-definable. In fact it given by a conjunction of pure probability logic formulas with parameters.*

Proof. Let (c_i) be a basis for the orthogonal complement of $L^2(\Xi)$ in $H_{\tilde{0}}$. We have $a \in X^*$ iff $(R(x, a), c_i) = 0$ for each i. By piecewise $\bigwedge$ interpretability of the Hilbert space structure, each map $a \mapsto (R(x, a), c_i)$ is (a uniform limit of) definable relations on X. So X^* is $\bigwedge$-definable.

To see that the relevant formulas use only expectation quantifiers, we may assume $\|R(a, z)\|_2 \leq 1$. Approximate c_i by continuous functions $c_{i,j} \in C(S_Z(\mathbb{U}))$, so that $\|c_i - c_{i,j}\|_{L^2(\Xi)} < 2^{-j}$. Then the condition $(R(x, a), c_i) = 0$ can be written as the conjunction of $(R(x,a), c_{i,j}) \leq 2^{-j}$. Now $(R(x,a), c_{i,j}) = E_z R(x,z) c_{i,j}(z)$ is visibly obtained by an expectation quantifier. □

(Is X^* $\bigwedge$-definable by pure probability logic formulas without parameters?)

Next we prove (b). This concerns only (X, Z, R), so Y, R' are not involved, and in the proof of Lemma 3.18 we will use the letter y for a second variable ranging over X.

Using the two projections from $S_{xz}(\mathbb{U})$, we can view both measure algebra of $\hat{X}$ and $\tilde{0}$ as subalgebras of the measure algebra $B_{xz}(\mathbb{U})$. Let $\tilde{0}[\hat{X}]$ denote the σ-subalgebra of $B_{xz}(\mathbb{U})$ they generate; this is just the measure algebra of the product measure space of $(Z, \tilde{0}) \times (\hat{X}, \nu)$.

Lemma 3.18. $\nu(X^*) = 1$.

Proof. We let x, y range over (X, ν), while z ranges over (Z, μ). Since μ, ν commute, we have, as our principal use of Fubini,

$$E_x E_y E_z (R(x,z) \wedge R(y,z)) = E_z E_x E_y (R(x,z) \wedge R(y,z)).$$

Now $E_x E_y (R(x,z) \wedge R(y,z)) = E_x R(x,z) E_y R(y,z) = (E_x R(x,z))^2$ by property (2) of probability quantifiers, so

$$E_x E_y E_z (R(x,z) \wedge R(y,z)) = E_z((E_x R(x,z))^2). \tag{1}$$

Let $\beta = E(R(x,z)|\tilde{0}[\hat{X}])$ be the conditional expectation of $R(x,z)$ (as an element of $L^2(S_{x,z}(\mathbb{U}))$) to an element of $L^2(\tilde{0}[\hat{X}])$.

We will express each side of (1) in terms of β. First, using Theorem 3.16(1), for any $a, b \in X$,

$$E(R(a,z) \wedge R(b,z)|\tilde{0}) = E(R(a,z) : \tilde{0}) E(R(b,z)|\tilde{0}) = \beta_a \beta_b$$

so

$$E_z(R(a,z) \wedge R(b,z)) = \int_{\tilde{0}} (E(R(a,z) : \tilde{0}) E(R(b,z)|\tilde{0}) = \int_{\tilde{0}} \beta_a \beta_b = \int_{\tilde{0}} \beta_{\hat{a}} \beta_{\hat{b}},$$

where $\int_{\tilde{0}}$ denotes the integral of a $\tilde{0}$-measurable function (note all our functions are bounded and so integrable). The last step expresses the fact that $\hat{a}$ determines $\operatorname{tp}(a/\tilde{0})$, and in particular $\beta_a = \beta_{\hat{a}}$ depends only on $\hat{a}$.

Note that $\int_{\hat{a} \in \hat{X}} \int_{\hat{b} \in \hat{X}} \beta_{\hat{a}} \beta_{\hat{b}} = \left(\int_{\hat{a} \in \hat{X}} \beta_{\hat{a}} \right)^2$. But $\int_{\hat{a} \in \hat{X}} \beta_{\hat{a}} = E(\beta|\tilde{0})$. Thus

$$\begin{aligned} E_x E_y E_z (R(x,z) \wedge R(y,z)) &= \int_{\hat{a} \in \hat{X}} \int_{\hat{b} \in \hat{X}} \int_{\tilde{0}} \beta_{\hat{a}} \beta_{\hat{b}} \\ &= \int_{\tilde{0}} E(\beta|\tilde{0})^2 = \|E(\beta|\tilde{0})\|_2^2, \end{aligned} \tag{2}$$

where the norm is taken in $H_{\tilde{0}}$. Here we used ordinary Fubini, allowed since the integrand is measurable for the product measure.

On the other hand, for $c \in Z$, $E_x R(x,c)$, as the value at c of a pure probability formula with respect to ν, depends only on the pure probability type of c over ν, and hence only on $\hat{c}$, the image of c in Ξ, which determines $\mathrm{tp}(c/\hat{X})$. (Here we use definability of ν over $\hat{X}$-parameters.) By factoring $E(\beta|\Xi)$ through $E(\beta|\hat{X} \times \Xi])$, we see that for almost all $\hat{c}$, $E_x R(x,\hat{c}) = E(\beta|\Xi)(\hat{c})$.

Thus, squaring this equality and integrating now over $\hat{c} \in \Xi$,

$$E_z((E_x R(x,z))^2) = \|E(\beta|\Xi)\|_2^2, \tag{3}$$

where now the norm is taken in $L^2(\Xi)$. By (1)–(3) we have

$$\|E(\beta|\tilde{0})\|_{L^2(\tilde{0})} = \|E(\beta|\Xi)\|_{L^2(\Xi)}.$$

Now $E(\beta|\Xi)$ is the orthogonal projection to $L^2(\Xi)$ of $E(\beta|\tilde{0}) \in H_{\tilde{0}}$. Since they have the same norm, we must have $E(\beta|\tilde{0}) \in L^2(\Xi)$.

Let ψ be any continuous "test function" on $\hat{X}$. Then $\psi\nu$ is another measure on X; it may not be 0-definable but it is $\hat{X}$-definable, so that the above result applies. We obtain $E(\psi\beta|\tilde{0}) \in L^2(\Xi)$. By factoring first through the product algebra $B_{\hat{X}} \times \tilde{0}$, it follows that for almost all $q \in \hat{X}$, $E(\beta_q) \in L^2(\Xi)$, concluding the proof. □

Proof of Theorem 3.16(2). It remains to prove (c). Let $a \in X^*$. Recall $\beta_a \in L^2(\tilde{0})$ is the conditional expectation of $R(a,z)$ relative to $\tilde{0}$, equivalently the orthogonal projection of $R(a,z) \in L^2(S_M(Z))$ to $L^2(\tilde{0})$; while α_a is the orthogonal projection of $R(a,z)$ to $L^2(\Xi)$; α_a depends only on $\mathrm{tp}(a/\Xi)$ and hence on $q = \hat{a}$. Similarly define β'_b, α'_b for R'. Then by Theorem 3.16(1),

$$E_z(R(a,z) \wedge R'(b,z)) = \int_{r \in S(\tilde{0})} \beta_a \beta'_b = (\beta_a, \beta'_b).$$

The last term is the inner product in $L^2(\tilde{0})$. By definition of X^*, we have $\beta_a = \alpha_a$, so

$$(\beta_a, \beta'_b) = (\alpha_a, \beta'_b) = (\alpha_a, \alpha'_b).$$

(The last equality is by the characteristic property of orthogonal projections to closed subspaces of Hilbert space, $(P(u), v) = (P(u), P(v))$.) Thus

$$E_z(R(a,z) \wedge R'(b,z)) = (\alpha_a, \alpha'_b),$$

where now the inner product is computed in $L^2(\Xi)$, and hence proves independence over Ξ. □

Remark 3.19. Assume L is countable.

(1) Aside from the sharper conclusion, Theorem 3.16 has a considerably wider domain of applicability than a purely L^2-based statement such as Theorem B.11, which applies only to a random 2-type. For example Theorem 3.16 applies when $\mathrm{tp}(a/\Xi) = \mathrm{tp}(b/\Xi)$, frequently an important situation, though $\mathrm{tp}(a, b)$ is certainly not random in this case. An example of this is given in Appendix C.

(2) Note X^* is defined in terms of (X, Z, μ, R) alone. It is shown to have full ν-measure for any commuting ν. And if $a \in X^*$, statistical independence over Ξ is proved for *any* $(Y, R' \subset Y \times Z)$.

(3) A variation: Let (Z, μ) be a stochastic sort. Let $a \in X(M)$, $b \in Y(M)$, with $a \perp b$. Assume X has a Ξ-definable measure ν commuting with μ, and concentrating on $\mathrm{tp}(a)$. Then $R(a, z)$, $R'(b, z)$ are statistically independent over Ξ. It suffices for ν to be Borel-definable, in the sense of [45]. We do not use self-commutation of ν! This is proved in the same way as Theorem 3.16(2), but more easily; in Lemma 3.18 integration over $\hat{X}$ becomes unnecessary, since only one strong type is involved.

(4) Let (Z_1, μ_1) and (Z_2, μ_2) be stochastic sorts. For a measure one set of types q_2 on Z^2, if $(a, b) \models q_2$ then $a \perp b$. Here "types" can be taken to be Φ-types, where Φ is the family of all stable probability logic formulas.

(5) Let (Z, μ) be a self-commuting stochastic sort, and $R_i \subset Z^2$ a definable binary relation. Then for almost all types q on Z^n, if $(a_1, \ldots, a_n) \models q$ then the events $R_i(a_i, z)$ $(i = 1, \ldots, n)$ are independent over $\tilde{0}$, and over Ξ in case μ is self-commuting. This follows inductively from (4) and Theorem 3.16, taking at first $X = X^{n-1}$, $Y = Z = X$ to obtain that $\bigwedge_{i \leq n-1} R(a_i, z)$ is statistically independent from $R(a_n, z)$ over Ξ.

(6) We used here the full Kim–Pillay space, without restricting the level of definability of the implied $\bigwedge$-definable equivalence relations. This is inevitable due to the starting data; our notion of independence uses the complete type of a and of b; in particular if a is 0-definable or lies in the bounded closure of 0, via a formula involving quantifiers, then $a \perp a$ holds. On the other hand the deduction of (2) from (1) uses probability quantifiers only. Since the Hilbert space is PPL interpretable, it should be possible to formulate a version of (1) and hence of the full theorem with definability in terms of probability quantifiers, given a stronger assumption of independence at the quantifier-free level.

3.20. ***Interpretative power of probability logic in a binary relational language.*** Let M be an L-structure with all sorts stochastic with commuting expectation quantifiers; for simplicity take a single sort X, and assume the measure on X is self-commuting.

Recall $\hat{X}$ is the biggest bounded quotient of X. If $f : \hat{X} \to \mathbb{R}$ is a continuous function, then $\alpha(x) = f(\hat{x})$ is an M-definable function. Let $M_{\hat{X}}$ be the result of adding to the language all such functions α, or equivalently, a countable subset that separates points. $M_{\hat{X}}$ is the expansion by all parametrically definable relations that are definable in M_A for *any* elementary submodel A of M, M_A being the expansion of M by constants for A. Moving to $M_{\hat{X}}$ is known as *working over the algebraic closure of the empty set*, here in the sense of continuous logic.

Assume that L consists of unary and binary relations, and possibly unary function symbols. Let L_{qf} denote the set of quantifier-free formulas, and L_{prob} the result of closing L_{qf} under continuous connectives and expectation quantifiers.

Corollary 3.21. *Assume L is binary, and countable. Away from a measure zero set of n-types, the L_{prob} type of a tuple $(a_1, \dots, a_n)$ of M is determined by the quantifier-free type of $(a_1, \dots, a_n)$ along with the values of **unary** $\hat{X}$-definable formulas $\alpha(a_i)$.*

Equivalently, for each $\phi \in L_{\mathrm{prob}}$ there exist formulas $\alpha_1, \alpha_2, \dots$ of $L_{\hat{X}}$, each taking a single variable from among $x_1, \dots, x_n$, L_{qf}-formulas $\beta_k(x_i, x_j)$, each taking two variables from among the x_i, and a Borel function Ψ, such that

$$E(|\phi - \Psi(\alpha_1, \alpha_2, \dots, \beta_1, \beta_2, \dots)|) = 0.$$

In other words, for almost all $x_1, \dots, x_n$, $\phi = \Psi(\alpha_1, \alpha_2, \dots, \beta_1, \beta_2, \dots)$.

Proof. Let us first see that the second statement follows from the first. Let Z_n be the space of L_{prob}-types on X^n, $Z = Z_n \times_{(Z_1)^n} \hat{X}^n$, W the space of Δ-types where Δ consists of all qf formulas along with $\hat{X}$-definable unary formulas. We have a natural restriction map $r : Z \to W$. By the first statement, there exists a measure-one set $Z' \subseteq Z$ such that r is injective on Z'. We may take Z' to be an F_σ set, i.e., $Z' = \bigcup_n Z_n$ is a countable union of compacts (seeing that Z is compact). Let $W_n = r(Z_n)$. Then W_n is a closed subset of W, and r^{-1} is continuous on W_n. It follows that r^{-1} is Borel on $\bigcup_n W_n$; and any continuous function ϕ on Z can be expressed as $\Psi(r(z))$, where $\Psi = \bigcup_n (r^{-1}|W_n)$.

Next let us prove that the L_{prob} type is indeed determined by the given data. The unary relations α arising from continuous functions on $\hat{X}$ can be recombined to give the map $x \mapsto \hat{X}$. So it suffices to show that for $\phi \in L_{\mathrm{prob}}$, the value $\phi(a_1, \dots, a_n)$ is determined a.e. by the quantifier-free type of $(a_1, \dots, a_n)$ along with the elements $\hat{a}_i \in \hat{X}$. Using Hoover's Theorem 3.6, we may take ϕ to have the form $E_w \psi(w, a_1, \dots, a_n)$, where w may be a tuple, and ψ is quantifier-free. As usual in quantifier-elimination, working inductively, we may assume w is a single variable. By Stone–Weierstrass we can take ψ to be a polynomial in basic formulas $R(w, x_j)$. Since E_w is additive, it suffices to determine the value of each monomial, i.e., of products of such basic relations. In the presence of function symbols, we view a relation $R'(fw, gx_j)$ simply as another relation $R''(w, x_j)$. We may collect together all relations belonging to a given variable x_j to obtain a single relation $R_j(w, x_j)$. The value $E_w R_j(w, a_j)$ is determined by $\hat{a_j}$. Finally the value of $E_w \prod_j R_j(w, x_j)$ is just the product of these last, by Remark 3.19(5). $\square$

Remark 3.22. (1) Here $\hat{X}$ should be viewed as a topological *structure*. The relations are these: for each stable qf formula $\phi(x, y)$, we have a map on $\hat{X}^2$ giving the generic value of ϕ at $(p, q) \in \hat{X}^2$.

(2) It would be interesting to determine when Ψ can be taken to be continuous and not just Borel. If one is content with quantifier-elimination up to 99%, rather than almost everywhere, Ψ can be taken to be a continuous function of finitely many variables: for each $\phi = \phi(x_1, \ldots, x_n) \in L_{\text{prob}}$, there exists a quantifier-free $L_{\hat{X}}$ formula ϕ' such that $E(|\phi - \phi'|) \leq 0.001$.

(3) A (real-valued) $\mathbb{U}$-definable formula ψ is a *matrix coefficient* if the set of $\mathrm{Aut}(\mathbb{U})$-conjugates of ψ spans a finite-dimensional space; equivalently, ψ factors through a definable map from X to a $\bigwedge$-interpretable finite-dimensional Hilbert space. In place of working over the algebraic closure, one can make similar statements in terms of matrix coefficients maps or in terms of definable maps into $\bigwedge$-interpretable finite-dimensional Hilbert spaces.

(4) We did not restrict the definability level of the unary maps α_i in Corollary 3.21. In case we are working over an elementary submodel, it suffices to take qf-definable ones. If the Galois group of $\hat{X}$ is trivial, one can take qf-definable maps over a saturated model M, with the property that they are invariant under $\mathrm{Aut}(M/\hat{X})$. In general it should be possible to describe the quotient of $\hat{X}$ we require using probability logic definable functions; we do not take it up here, but see Remark 3.19(6).

This can be read as saying that with pure probability logic, over a binary language,[4] interesting finite or finite-dimensional structures are interpretable along with a map from M into them; and given these, nothing else can be interpreted that is not visible at the level of basic relation symbols.

The fundamental problem here is to extend the theory Theorem 3.16 to 4-amalgamation and higher. The following weak version would already be useful. Recall that in the presence of a notion of independence of two substructures over a third, an *independent system* of substructures is a family $(A_u : u \in S)$, where S is a simplicial complex, such that A_u is independent from $\bigcup\{A_v : \neg u \leq v\}$ over $\bigcup\{A_w : w < u\}$.

Problem 3.23. Assume μ is a strictly definable measure on X. Does there exist a canonical piecewise-interpretable independent system of measure algebras $(S_{\tilde{u}} : u \subset [n])$ containing the measure algebras $F(u)$ of formulas in variables from u?

[4] I.e., the signature has only binary relation and unary function symbols.

Part of the above statement is *existence* of such an independent system. This should be possible essentially using the result over a model M (Theorem B.11), but replacing M by a probability space of possible interactions with the variables; this only provides a highly "almost everywhere" result.

By Theorem B.8, at least a measure stationarity is obtained assuming higher amalgamation. For stable theories, the expansion required to obtain higher amalgamation is understood, see [42]. Could this be combined with stability of the measure algebras so as to give a more precise construction bringing out the geometry?

3.24. *Stability and NIP.* The following proposition — for stability and NIP — is a very special case of a powerful general theory of *randomization*, due to Ben Yaacov and Keisler. The proof we give for all three is a simple application of the Vapnik–Chervonenkis uniform law of large numbers.

Proposition 3.25. *Stability, NIP and pNIP are preserved by probability quantifiers: assume $\psi(u, x; y)$ is stable (resp. NIP, pNIP). Then $(E_u)\psi(u, x, y)$ is stable (resp. NIP, pNIP).*

Proof. Suppose $(E_u)\psi(u, x, y)$ is unstable. Then there exist $\alpha < \beta \in \mathbb{R}$ and (a_i, b_i) ($i \in \mathbb{N}$) such that $(E_u)\psi(u, a_i, b_j) < \alpha$ when $i < j$ while $(E_u)\psi(u, a_i, b_j) > \beta$ when $i > j$. By [74], for some N there exist $c_1, \dots, c_N$ such that for any a, b,

$$\left|(E_u)\psi(u, a, b) - \frac{1}{N}\sum_{k=1}^{N}\psi(c_k, a, b)\right| < \frac{1}{3}(\beta - \alpha).$$

Let $\alpha' = \alpha + \frac{1}{3}(\beta - \alpha)$ and $\beta' = \beta - \frac{1}{3}(\beta - \alpha)$. By refining the sequence we may assume $\lim_{j\to\infty}\lim_{i\to\infty}\psi(c_k, a_i, b_j) = \gamma_k$ and $\lim_{i\to\infty}\lim_{j\to\infty}\psi(c_k, a_i, b_j) = \gamma_k'$ both exist. Now for $i < j$ we have $\frac{1}{N}\sum_{k=1}^{N}\psi(c_k, a_i, b_j) < \alpha'$ while for $i > j$ rather $\frac{1}{N}\sum_{k=1}^{N}\psi(c_k, a_i, b_j) > \beta'$. Thus

$$\frac{1}{N}\sum_{k=1}^{N}\gamma_k' < \frac{1}{N}\sum_{k=1}^{N}\gamma_k.$$

But by stability of $\psi(c_k, x, y)$ we have $\gamma_k = \gamma_k'$; a contradiction.

A similar proof works for NIP, once we know that the value of the sample size N in the Vapnik–Chervonenkis theorem can be bounded polynomially; in the case of pNIP, we need the bound to depend polynomially on both the desired approximation and on the pNIP degree (equivalently, on the Vapnik–Chervonenkis dimension).

To simplify notation take the special case of a $\{0, 1\}$-valued relation $\psi(u, x, y)$. Let d' be the Vapnik–Chervonenkis dimension of $\psi(u; x, y) \wedge \psi(u; x', y')$ viewed as a relation between u and x, y, x', y'. Let d be the Vapnik–Chervonenkis dimension of $\psi(u, x, y)$ viewed as a relation between x and u, y.

Let $n, m \in \mathbb{N}$; let B be a set of size m (in the y-sort), and let $a_1, \ldots, a_r$ have distinct $(Eu)\psi(u; x, y)$ types over A, at resolution $1/n$; in other words if $i \neq j$ then for some $b \in B$, $|(Eu)\psi(u; a_i, b) - (Eu)\psi(u; a_j, b)| > 1/n$. In particular (possibly after interchanging i, j) we have $\mu(\{u : \psi(u; a_i, b) \wedge \neg\psi(u; a_j, b)\}) > 1/2n$. We have to bound r polynomially in m, n.

By Proposition 2.18, there exist a set C of size $\leq (16d'n)^2$ such that for any $i \neq j$, for some $b \in B$ and some $c \in C$ we have $\psi(c; a_i, b) \wedge \neg\psi(c; a_j, b)$ (or vice versa). Thus the elements a_i have distinct ψ-types over $B \cup C$. By assumption, if $|B| \geq d$ we have $r \leq (|B| + |C|)^d$; this gives the required polynomial bound. □

From the above (either using Hoover's normal form, or induction on complexity of the formula) we obtain:

Corollary 3.26. *Let L be a language, possibly of continuous logic. Let M be an L-structure. Assume each basic formula is stable under* $\mathrm{Th}(M)$, *with respect to any partition of the variables into two nonempty sets. Then every pure probability logic formula is stable.*

A metric space is said to have *finite packing dimension* if for some $C, \alpha > 0$, for all sufficiently large n, any set of disjoint balls of radius $1/n$ has size at most Cn^α. The following is Theorem 4.1(c) of [60].

Proposition 3.27 (Lovász–Szegedy). *Let $\phi(x, y)$ be a* $\{0, 1\}$*-valued NIP formula on $X \times Y$.*

Assume given an invariant, generically stable measure $\mu(y)$, with associated expectation operator E_y. Define a premetric d on X by

$$d(a, b) = E_y(|\phi(a, y) - \phi(b, y)|) = \mu(\phi(a, y) \triangle \phi(b, y)).$$

Let M be a model, and $\overline{M}$ the completion of $X(M)$. Then $\overline{M}$ has finite packing dimension, depending only on the Vapnik–Chervonenkis dimension of ϕ.

The same is true for the L^2-distance $d_2(a, b) = \mu(\phi(a, y) \triangle \phi(b, y))^{1/2}$.

Proof. Let δ be the Vapnik–Chervonenkis dimension of ϕ. By the Sauer–Shelah lemma, the number of ϕ-types over an N-element set is bounded by $O(N^\delta)$. Assume the $1/n$-balls around $a_1, \ldots, a_k$ are disjoint. We have to bound k polynomially in n. For $i \neq j$ we have $d(a_i, a_j) \geq 1/n$, so the measure of either $\phi(x, a_i) \wedge \neg\phi(x, a_j)$ or the dual set is $\geq 1/2n$. Let $N = 16\delta n \log(16\delta n)$ and let $c_1, \ldots, c_N$ be as in Proposition 2.18. Then for each $i \neq j$ for some $\nu \leq N$ we have $\phi(a_i, c_\nu) \wedge \neg\phi(a_j, c_\nu)$ or vice versa. Thus the a_i have distinct ϕ-types over $c_1, \ldots, c_N$. The number of such types is at most $O(N^\delta)$. So $k \leq O(N^\delta) \leq O((n \log(n))^\delta)$.

If we use d_2 then $d_2(a_i, a_j) \geq 1/n$ implies $d(a_i, a_j) \geq 1/n^2$, so the same argument gives $k \leq O((n^2)^{2\delta}) = O(n^{4\delta})$. □

Remark 3.28. The proof of Proposition 3.27 is valid for any definable measure μ on ϕ-types, using Remark 2.19. Since the measure is definable, it suffices to consider $a_1, \dots, a_k$ in a model M; while $c_1, \dots, c_N$ may be taken in an elementary extension M^*.

3.29. ***A categoricity theorem, following Gromov, Vershik, Keisler.*** We now formulate a uniqueness theorem for probability logic structures carrying a metric and a definable measure of full support. For compact measure spaces, this is a theorem of Gromov's; Vershik [75] gave a simpler proof. All compact structures have categorical continuous logic theories; the point here is that only expectation quantifiers are used.

The result also bears a close relation with the uniqueness theorems for pseudofinite structures of Keisler [52, p. 34; 53, 3.2.9]; but note the strong property B4 assumed there, and not necessarily valid in our setting, e.g., for the random graph. (It is valid however when the measure is on the model itself, as is the case in Gromov's theorem.)

In our application to approximately homogeneous approximate equivalence relations, the theory itself ensures full support, i.e., when the volume of a ball of a given radius $r > 0$ is bounded above 0.

We prove the theorem without uniform full support, compactness or σ-additivity assumptions. In this case the result may be thought of as a probability logic analogue to uniqueness theorems for prime models, rather than a categoricity theorem. Note that it gives in particular a "soft" proof of the Gromov–Vershik theorem, different from Vershik's, using a basic model-theoretic "preservation theorem": if the universal theory of M contains that of N, then M embeds into an elementary extension of N.

Let L be a continuous logic language; it has in particular a formula $d(x, y)$ for a metric, and various additional real-valued relations, uniformly continuous with respect to the metric. Adjoin expectation operators, and let T be a pure probability logic theory of L; thus we have a class C of formulas ϕ including all quantifier-free formulas, and closed under expectation operators. We say T is ppl-complete if for every $\phi(x) \in C$, T determines $(Ex)\phi$.

Let $M \models T$. Recall that the expectation quantifiers induce a measure on the type space $S_x(M)$, so that any $\bigwedge$-definable set over M is assigned a measure. M is said to have *full support* if the measure of any ball is positive. M is *complete* if it is complete as a metric space.

Theorem 3.30. *Let T be a complete theory of pure probability logic. If M, N are two complete models of T with full support, then $M \cong N$. Moreover, any two tuples in M with the same pure probability logic type are conjugate by an automorphism of M.*

Proof. Let us view every formula using connectives and expectation quantifiers (the class C above) as basic. Write $M \leq N$ to mean that any basic formula θ satisfies $\theta^N(a) = \theta^M(a)$ whenever $a \in M^n$.

Claim 1. *Let M_1, M_2 be two complete models of T with full support. Then the universal theories of M_1, M_2 are equal.*

Proof. Let $\phi(x)$ be a basic formula of L, where $x = x_1, \dots, x_n$. It suffices to show that if $\phi(a)^{M_1} \geq 0$ for all $a \in {M_1}^n$, then $\phi(b)^{M_2} \geq 0$ for all $b \in {M_2}^n$. Suppose for contradiction that $\phi(b)^{M_2} < 0$. By (uniform) continuity, for some $\epsilon > 0$, for any $b' \in \prod_{i=1}^n B_\epsilon(b_i)$, we have $\phi(b') < 0$. By the full support assumption the measure of each of the balls $B_\epsilon(b_i)$ is nonzero; thus the same is true of their product. Let $\psi = \min(0, \phi)$. Then $E_x \psi^{M_2} < 0$. But clearly $E_x \psi^{M_1} \geq 0$, contradicting the assumption that the pure probability theories are the same. $\square$

Claim 2. *Let $M \leq N$ with M complete, and $c \in N \setminus M$. Then for some $\epsilon > 0$, the ball $B_\epsilon(c)$ is disjoint from M.*

Proof. If there were no such ϵ, we could find a sequence of elements of M approaching c; but M is complete, so $c \in M$ would follow. $\square$

Claim 3. *Let $M \leq N$. Let $B = B_\epsilon(c)$ be a ball in N with no points in M. Then $\mu(B) = 0$.*

Proof. Let $\epsilon' = \mu(B)$. Find in M elements $a_1, \dots, a_k$ such that $\beta := \mu\big(\bigcup_{i=1}^k B_\epsilon(a_i)\big)$ is as large as possible, to within ϵ', so that the union of $k+1$ ϵ-balls of M has volume $< \epsilon' + \beta$. By Claim 1, the same is true in N. However, $\mu\big(\bigcup_{i=1}^k B_\epsilon(a_i) \cup B_\epsilon(c)\big) = \beta + \epsilon'$, a contradiction. $\square$

Let us now prove the theorem. By Claim 1, M, N have the same universal theory; so N embeds into an elementary extension M^* of M; we view it as so embedded. Let $c \in N$. By Claim 2, if $c \notin M$ then some ball $B_\epsilon(c)$ is disjoint from M, and by Claim 3, $\mu(B_\epsilon(c)) = 0$. But this contradicts the full support assumption on N. Thus $N \subseteq M$. Similarly, $M \subseteq N$, so $M = N$ and in particular $M \cong N$.

For the "moreover", if a', a'' have the same type, enrich M by additional real-valued relations $\phi(x, a')$ (respectively $\phi(x, a'')$), for ϕ a probability logic formula, to obtain structures M', M'' with the same pure probability logic theory, and with full support. By the main part of the theorem, there exists an isomorphism $M' \to M''$, hence an automorphism of M with $a' \mapsto a''$. $\square$

Remark 3.31. The statement and proof of Theorem 3.30 remain valid for many-sorted theories. Each sort is assumed to be endowed with a metric, and with expectation quantifiers; M and N are assumed to be complete and of full support in each sort separately.

3.32. ***Local probability logic.*** We require a slight variant, *local probability logic*.

We work with local continuous real-valued logic as in Section 2.3. Recall that a local relation $\phi(x_1, \dots, x_n)$ has bounded support, determined by ρ^* and some compactly supported continuous function $b = b_\phi : \mathbb{R} \to \mathbb{R}$; we guarantee that

$$|\phi(x_1, \dots, x_n)| \le b_\phi(\max_{i,j} \rho^*(x_i, x_j)).$$

Our description will depend in addition on a choice of positive reals $C_1 \le C_2 \le \cdots$; C_k should be thought of as a bound for the measure of a ρ^*-ball of radius $2k$.

Given a formula $\phi(x, y_1, \dots, y_n)$ with $n + 1$ variables, $n \ge 1$, we allow an expectation quantifier, so that we can form $(Ex)\phi(x, y_1, \dots, y_n)$.

By a *probability* or *expectation quantifier in x* we mean a syntactical operation from formulas $\phi(x, y)$ (with y a nonempty sequence of variables distinct from x) to formulas $E_x\phi$ in the variables y, satisfying (2–4) of Section 3, and this generalization of (1):

$(1_{\mathrm{loc},k})$ For any continuous $\beta : \mathbb{R} \to [0, 1]$ supported on $[-k, k]$,

$$E_x\beta(\rho^*(x, y)) \le C_k.$$

The numbers C_k are also used in the inductive definition of the syntactic bound for the modulus of continuity of a formula; namely the modulus of $E_x\phi(x, y)$ is $C_{b_\phi+1}$ times the modulus of continuity of $\phi(x, y)$.

In practice, we concentrate on the case where the locality relation is induced by a two-valued relation R; namely $\rho^* = d_R$. Theorem 4.3 will be formulated in this setting (though it could be generalized). The idea is that quantification and expectation can only be taken within d_R-balls of some bounded radius.

3.33. ***Semantics.*** A model M for local probability logic is a model M for the underlying local continuous logic theory, along with a definable measure on the type space $S_x(M)$, such that for any local formula $\phi(x, y)$, we have $\int \phi(x, b) = (E_x\phi)(b)$. In particular, the measure of a d_R-ball of radius k is at most C_k.

In locally pseudofinite semantics, we begin with a family of locally finite graphs G_i, letting ρ^* be the graph distance, and using a multiple $\mu = c_i\mu_{\mathrm{count}}$ of the counting measure; such that the volume of a ball of radius 1 is at most C_1.

By *pure* (local) probability logic we mean the local formulas obtained from the basic ones using local connectives (Section A.1) and expectation quantifiers E_z alone. Due to the locality stipulation in the formation of formulas, there may be no pure probability logic formulas without free variables. We define the pure probability logic *theory* of a structure M by allowing universal quantifiers on the left. Thus to give this theory is equivalent to determining the closure of $\phi(M)$ for any tuple ϕ of formulas. Of course once a constant is added sentences do appear; in the proposition below, where a constant is assumed, the theory can be taken to be the set of values of *sentences*. By Claim 1 of Theorem 3.30, the additional universal quantifiers do not add information here, when M has full support.

Proposition 3.34. *Let (X, a) be a complete pointed model of a local probability logic theory, such that any ball has finite, nonzero measure. Then the isomorphism type of (X, a) is uniquely determined by the probability logic theory of (X, a). In other words if (Y, b) is another structure with the same properties, and the type of a in X equals the type of b in Y, then $(X, a) \cong (Y, b)$.*

Similarly for k-pointed structures.

Proof. Note that Theorem 3.30 applies to each closed neighborhood of the distinguished point. The proof is the same as of Theorem 3.30, but to begin with choose a type q in variables $x_{i,j}$, $i, j \in \mathbb{N}$, such that $x_{k,j}$ lies at distance $\leq k$ from a; with q random in the product space of the balls of radii $1, 2, 3, \ldots$ around a. Use also the additional relations allowed there of the form $\phi(a, x)$, for ϕ a probability logic formula.

The k-pointed case follows from the 1-pointed case, as the language may include constants. Compactness of a metric space implies separability and completeness and so only strengthens the hypothesis. □

Remark 3.35. We will obtain X as the completion of a (locally) saturated probability logic structure M, with respect to a definable pseudometric d. This includes a quotient with respect to the equivalence relation $d(x, y) = 0$, which is assumed (locally) cobounded (this is equivalent to the (local) compactness assumption on X). Let P be a 1-type of M with respect to pure probability logic, and let $\bar{P}$ be the image of P in X. Then Proposition 3.34 assures us that the (isometric) isomorphism group G of X is transitive on $\bar{P}$. *This does not, in itself, mean that* $\mathrm{Aut}(M)$ *is transitive on* P, since the induced map $\mathrm{Aut}(M) \to \mathrm{Aut}(X)$ may not be surjective; the pure probability logic type may not generate a complete type.

If we use full continuous logic, including the expectation quantifiers, we can enrich X by predicates for all the images on X of 0-definable relations on M. They are all closed in the logic topology, and hence in the metric topology. Also take M is $|L|^+$-saturated and homogeneous. In this case, the natural map $G \to \mathrm{Aut}(X)$ is surjective. To see this, let $\bar{a}$ be a random sequence from X as in the proof of Theorem 3.30, and $\bar{b} = g(a)$. Lift $\bar{a}$ to $a \in M$. Then $\bar{b}$ lifts to $b \in M$ satisfying the same type. By saturation there exists an automorphism of M taking a to b. Proposition 3.34 is similar.

Definition 3.36. A sequence of finite graphs (Ω_n, R_n) is *approximately homogeneous* if for any pure probability formula in one variable $\phi(x)$, the value of ϕ becomes constant as $n \to \infty$:

$$\lim_{n,n' \to \infty} \sup_{x \in \Omega_n, x' \in \Omega_{n'}} |\phi(x) - \phi(y)| = 0.$$

A similar definition applies in local probability logic

The sequence is *approximately homogeneous a.e.* if any pure probability formula in one variable $\phi(x)$, for some $v = v(\phi)$, for any $\epsilon > 0$, for all sufficiently large n we have

$$\mu(\{x \in \Omega_n : |v - \phi(x)| > \epsilon\}) < \epsilon.$$

In local probability logic, the sequence Ω_n is approximately homogeneous a.e. if for any local pure probability formula in one variable $\phi(x)$, for some $v = v(\phi)$, for any $\epsilon > 0$ and m, for all sufficiently large n and any ball B of Ω_n of radius m,

$$\mu(\{x \in B : |v - \phi(x)| > \epsilon\}) < \epsilon.$$

Equivalently, for any continuous $\beta : \mathbb{R} \to [0, 1]$ with compact support, for all sufficiently large n we have

$$(E_x \beta(\rho^*(x, t))|v - \phi(x)|)^{\Omega_n} < \epsilon.$$

Remark 3.37. Let us formulate the notion of a sequence of graphs approaching a 1-homogeneous graph in probability logic, in terms used in combinatorics ([69; 60]; compare also [11]).

In particular the measure of the set of neighbors $R(a) = \{b : (a, b) \in R\}$ approaches some real number ϖ. Let N be the set of connected graphs on $m + 1$ vertices. Given $a \in \Omega$, and $\gamma \in N$, let $C(\gamma, a)$ be the set of graph embeddings $\gamma \to \Omega$ with $0 \mapsto a$. Define the *local statistics function* $\mathrm{LS}_m : \Omega \to [0, 1]^N$ by

$$\mathrm{LS}_m(a)(\gamma) = \mu_m(C(\gamma, a)) = |C(\gamma, a)|/\varpi^m.$$

Say (Ω, R) is (m, ϵ)-*homogeneous* if the range of LS_m is concentrated in an ϵ-ball (for sup metric on $\mathbb{R}^N$). If (Ω, R) and (Ω', R') are both (m, ϵ)-homogeneous, we say that they are (m, ϵ)-close if the respective ranges intersect.

4. Stabilizer theorem for approximate equivalence relations

Two metrics d, d' are *commensurable at scale* α if an α-ball of d' is contained in finitely many α-balls of d, and vice versa; k-commensurable at scale α if the number of balls needed is $\leq k$.

A metric space is k*-doubling at scale* α if $d, \frac{1}{2}d$ are k-commensurable at scale α.

Definition 4.1. Let $\Gamma = (\Omega, R)$, where R is a symmetric, reflexive binary relation. R is a k*-approximate equivalence relation* if condition (1) holds. It is a *near equivalence relation* if for some finitely additive measure μ on Ω, (2,3) hold. R is *an amenable approximate equivalence relation* if (1–3) hold.

(1) (Main axiom; "doubling") For all a, a 2-ball $R^2(a)$ is a union of at most k 1-balls $R(b)$.

(2) For some $\varpi > 0$ and $\kappa > 0$, for all $a \in \Omega$, $(1/\varpi) \leq \mu(R(a)) \leq \varpi$; and $\mu(R^3(a)) \leq \kappa$.

(3) (Weak Fubini) For some $\vartheta > 0$, for all a, $\mu(\{b : \mu(R(a) \cap R(b)) \geq \vartheta\}) > \vartheta$.

Given (2), weak Fubini follows from Fubini applied to $\{(x, y) : R(a, x) \wedge R(x, y)\}$, a subset of $R(a) \times R^2(a)$ of measure at least equal to $1/\varpi^2$. We note in the lemma below that it automatically holds (assuming 2) if R is replaced by the distance-two relation R^2; or given (1,2), for (Γ, R) with transitive automorphism group.

Lemma 4.2. (1) *If* (Ω, R) *satisfies* (2) *then* (Ω, R^2) *is an amenable approximate equivalence relation.*

(2) *If* (Ω, R) *satisfies* (1,2) *and has a transitive automorphism group, then* (Ω, R) *is an amenable approximate equivalence relation.*

Proof. (1) We use the graph analogue of "Rusza's trick". Namely, a maximal disjoint set of balls $R(a_i)$ contained in $R^3(a)$ must have size at most $\kappa\varpi < \infty$. By maximality, for any $b \in R^3(a)$ we have $R(b) \cap R(a_i) \neq \varnothing$ for some i, so $b \in R^2(a_i)$; thus $R^3(a) \subset \bigcup_i R^2(a_i)$, i.e., $R^3(a)$ is a union of at most $\kappa\varpi$ two-balls. From this, inductively, $R^{2+m}(a)$ is the union of $(\kappa\varpi)^m$ two-balls. In particular taking $m = 2$, we obtain Definition 4.1(1) for R^2. For (2) we use the same measure; since every two-ball contains a one-ball we have the lower bound upon $R^2(a)$; and since every three-ball of R^2 is contained in at most $(\kappa\varpi)^4$ two-balls, we have the upper bound on 3-balls of R^2. Finally to check (3), if $b \in R(a)$ then $R(a) \subset R^2(b)$ so $\mu(R^2(a) \cap R^2(b)) \geq \mu(R(a)) \geq (1/\varpi)$. This shows that (Ω, R^2) is an amenable approximate equivalence relation.

(2) Note that (1,2) imply that (3) holds for *some* a: if $R^2(a) = \bigcup_{i=1}^k R(a_i)$, then (3) must hold for some a_i, since for any $b \in R(a)$ we have $\mu(R(b) \cap R^2(a)) > 0$. Hence (3) holds for all a if we have homogeneity. □

Note that subspaces of Euclidean space are doubling at every scale. Let Ω be an ϵ-sphere packing of $\mathbb{R}^n$, i.e., a maximal set (of "centers") such that any two are at distance at least ϵ. So any 2ϵ-ball contains at least one point of Ω. It follows that (Ω, R) is a k-approximate equivalence relation, for appropriate k on the order of 2^n; where R is the "distance at most 1" relation.

If (Ω, R) is given to us but not $\mathbb{R}^n$, can we recover the relations corresponding to radius one half balls, or smaller balls? Theorem 4.3 gives an affirmative result in this direction. We begin with d_R which makes sense for distances ≥ 1, deduce ρ which is meaningful in distances between 0 and 1, and show that we still have some doubling, and ρ, d_R more or less fit together at the scale 1.

By definition, a set Z has measure zero iff $B \cap Z$ has measure 0, for all (finite measure) balls $B \subset \Omega$.

Let R be an amenable k-approximate equivalence relation on Ω. We obtain a countably additive measure on the type space, or on a sufficiently saturated elementary extension of (Ω, R). Let Φ_0 be the set of pairs $\phi(x)$, (α, β) where ϕ is a formula of pure probability logic in one variable, and (α, β) is a rational interval in $\mathbb{R}$ (i.e., $\alpha < \beta \in \mathbb{Q}$) *such that* $S(\phi, \alpha, \beta) := \{x : \alpha < \phi < \beta\}$ has measure 0; equivalently, in terms of expectations, $E_x C(\phi(x)) = 0$ for any continuous function C supported on (α, β). Let Ω_0 be the union of all $S(\phi, \alpha, \beta)$, with $(\phi, (\alpha, \beta)) \in \Phi_0$. This is a countable union, so Ω_0 has measure 0. Note that allowing $\phi'(x)$, (α, β), where ϕ' is a uniform limit of formulas, would not change this union. Let Ω^* be the complement of Ω_0. Then Ω^* is a partial type of pure probability logic, has full measure, and is the smallest such.

Theorem 4.3. *Let R be an amenable k-approximate equivalence relation on Ω. Assume the graph (Ω, R) is connected. Then there exists a formula $d(x, y)$ of local probability logic, without parameters, such that*:

(1) *d defines a pseudometric. d and d_R are k'-commensurable at scale* 1, *where k' depends only on k. d is k''-doubling at any scale $s \leq \frac{1}{2}$, where k'' depends only on k and s.*

(2) *In the completion of* (Ω, d) (*modulo* $d(x, y) = 0$), *all closed balls of radius ≤ 1 are compact. The images in the completion of all d_R-balls any radius are thus compact.*

(3) *For any $m \in \mathbb{N}$, let S_m be the distance $\leq 1/m$-graph of d. Then $S_m^m \subset R^4$, and*

(4) *for some $C = C_m > 0$, for all $a \in \Omega^*$ we have $\mu S_m(a) \geq C \mu R(a)$.*[5]

Proof. (1) Define $d_0(x, y)$ by the expression

$$E_z(|E_t(R(t, x) \wedge R(t, z)) - E_t(R(t, y) \wedge R(t, z))|).$$

The expectation quantifiers E_t are clearly local. The quantities $E_t(R(t, x) \wedge R(t, z))$ and $E_t(R(t, y) \wedge R(t, z))|$ both vanish unless $d_R(x, z) \leq 2$ or $d_R(y, z) \leq 2$; thus the E_z is also a legitimate local probability logic quantifier.

It is clear that d_0 defines a pseudometric.

By Proposition 3.8, $\psi(x, z) = E_t(R(t, y) \wedge R(t, z))$ is a stable formula; we will use this below. We remark that by the same argument, $d(x, y)$ is stable.

Let ϑ' be as in the weak Fubini axiom Definition 4.1(3), and choose $0 < \vartheta < \vartheta'$. Define $d = d_0/\vartheta$.

Claim 1. $d(x, y) \leq 1$ *implies* $d_R(x, y) \leq 4$.

[5]Without assuming weak Fubini, one obtains the same theorem but with $S_m^m \subset R^8$ at worse; it suffices to replace R by R^2, in Claim 1, using Lemma 4.2.

Proof. Pick z with $E_t(R(t,x) \wedge R(t,z)) \geq \vartheta'$. If $d_R(x,y) > 4$, we have

$$E_t(R(t,y) \wedge R(t,z)) = 0,$$

so $E_z(|E_t(R(t,x) \wedge R(t,z)) - E_t(R(t,y) \wedge R(t,z))|) \geq \vartheta'$. This shows that $d_0(x,y) < \vartheta'$ implies $d_R(x,y) \leq 4$. Since ϑ was slightly decreased from ϑ', we obtain the stated version with the weak inequality. □

So a 1-ball of d is contained in a 4-ball of d_R, and hence also in finitely many 1-balls of d_R.

Claim 2. *For any $s > 0$, a 1-ball of R is covered by a bounded number of s-balls of d. Moreover the bound depends only on s and k.*

Proof. Otherwise, in an ultraproduct, we have a 1-ball of R containing an infinite $s/2$-discrete subset for d; in particular there exist b, c with $d(b,c) \geq s/2$ and with the same Kim–Pillay type.

In the same ultraproduct, consider an element a avoiding any given countable set of b, c-definable measure zero formulas; in particular it does not divide over b or c. By Theorem 2.10 for such b, c, for any such random a we have

$$E_t(R(t,a) \wedge R(t,b)) = E_t(R(t,a) \wedge R(t,c)).$$

So $d(b,c) = 0$, a contradiction. □

Putting together Claims 1 and 2, we see that d, d_R are commensurable at scale 1. Moreover, a 1-ball of d is contained in a 4-ball of d_R and hence in k^3 1-balls of d_R; each of these is contained in a bounded number of s-balls of d, by Claim 2; so certainly a $2s$-ball of d is contained in a bounded number of s-balls of d, if $s \leq \frac{1}{2}$. This proves (1).

(2) Follows from the total boundedness of the balls of d (Claim 2), and completeness.

(3) Clear.

(4) Recall Ω^* from just above the theorem.

Claim 3. *For any $m > 1$ and any $c \in \Omega^*$, $\mu S_m(c) > 0$.*

Proof. We may replace (Ω, R) by an extension saturated for local formulas. Let $c \in \Omega^*$, and suppose for contradiction that $\mu S_m(c) = 0$. Let $P = \{c' : \mu S_m(c') = 0\}$. Since $c \in \Omega^*$, for some R-ball $B = R^l(a)$, $R^l(a) \cap P$ is not contained in a $\bigvee$-definable set of measure zero. Let $A = \{a_i : i \in I\}$ be a maximal subset of $P \cap R^l(a)$ such that the $S_{2m+2}(a_i)$ are disjoint, $a_i \in A$. Then the d-balls $S_{m+1}(a_i)$ cover $P \cap R^l(a)$. A cannot be infinite; otherwise, by Claim 2, infinitely many elements a_i of A are contained in a single $1/(2m+2)$-ball of d, say around x_0; but then x_0 lies in each of the $S_{2m+2}(a_i)$, contradicting their disjointness. Thus $A = \{a_1, \ldots, a_\nu\}$ is a finite set, and the d-balls $S_{m+1}(a_i)$ cover $P \cap R^l(a)$.

By saturation, for some $\epsilon > 0$, for all $c' \in R^l(a)$, $\mu S_m(c') < \epsilon$ implies that $c \in S_m(a_i)$ for some $i \leq \nu$. Since $\mu S_m(c') < \epsilon$ is a $\bigvee$-definable set, it cannot have measure 0 in $R^l(a)$; so some $S_m(a_i)$ has measure > 0. This contradicts $a_i \in P$. □

We have shown that $\mu S_m(c) > 0$ for $c \in \Omega^*$. By compactness, $\mu S_m(c)$ must be bounded strictly above 0, uniformly for all $c \in \Omega^*$. □

Remark 4.4. Fix $\epsilon > 0$, and define an ϵ-slice to be a $\bigvee$-definable set Z such that for any ball $R(a)$ we have $\mu(Z \cap R(a)) < \epsilon\mu(R(a))$.

Then for some $C = C_{m,\epsilon}$, for all a away from an ϵ-slice, $\mu S_m(a) \geq C\mu(R(a))$. Moreover, $C_{m,\epsilon}$ depends on m, ϵ, k alone. This follows from Theorem 4.3(4) by a compactness argument.

Question 4.5. While d is definable from R, almost contained in the distance-two relation R^2 and intuitively much finer, it is not clear under what conditions the image of R^2 in the completion of d has the same probability logic theory as (M, R^2). When does the image of R^2, and similar definable sets, have boundary of measure zero?

Remark 4.6. Initially, the stability-theoretic proof produced an $\bigwedge$-definable equivalence relation $S = \bigcap_m S_m$ corresponding to $d(x, y) = 0$, implying the existence of the relations $d(x, y) \leq 2^{-m}$ with explicitly exhibiting them. However, an analysis of the proof (in the case of ideals arising from a measure) shows that $(x, y) \in S$ iff for almost every z, $\mu(R(x) \cap R(z)) = \mu(R(y) \cap R(z))$. Thus we can take $S = \bigcap_n S_n$ with

$$x S_n y \iff \mu\{z : |\mu(R(x) \cap R(z)) - \mu(R(y) \cap R(z))| \geq 2^{-n}\} \leq 2^{-n}.$$

Putting this into real-valued based probability logic naturally leads to the smoother form used above. Note that a bounded L^1 function f on a finite probability space has small L^1-norm iff it is small in the sense of convergence in measure, i.e., for a small ϵ it takes value $> \epsilon$ only on a set of measure ϵ.

Example 4.7. Let G be a group, and X an approximate subgroup. Define $R(x, y)$ iff $x^{-1}y \in X$. Then R is an approximate equivalence relation. G acts on (G, R) on the left, by automorphisms. Since $d(x, y)$ of Theorem 4.3 is definable without parameters, G also preserves $d(x, y)$; it follows that $d(x, y) = 0 \iff x^{-1}y \in S$ for some S. This recovers the stabilizer theorem of [43].

4.8. ***Comparison to Lovász–Szegedy.*** After talking about this material in the Groups and Words meeting in Jerusalem in 2012, Nati Linial pointed out the relation to [60]. indeed, their definition of a *graphon* uses precisely the definition of d in Theorem 4.3. (Strictly speaking, they work in the case of a finite measure, or doubling constant one, whereas ours is only locally finite.) Moreover a more recent paper [61] concerns automorphisms of graphons so the overlap must be very considerable. From a model-theoretic viewpoint, the *graphon* is isomorphic to a quotient of the space of compact Lascar types of a theory of graphs, namely the canonical quotient topologized by formulas $\phi(x, b)$ where ϕ is the graph edge relation. In the presence of probability quantifiers, the KP space, and hence this quotient, carry a canonical measure.

Model-theoretically, one constructs first a saturated model, then a type space, the Lascar compact quotient, and (in the presence of probability quantifiers) measure on it. The graphon approach, by contrast, begins with the measure theory; but it is still able to construct parallels of the above objects. The proof that these objects are in fact the same, insofar as pure probability logic goes, requires the independence theorem Theorem 3.16.

5. Approximately symmetric approximate equivalence relations, and Riemannian models

Definition 5.1. (1) A *Riemannian homogeneous space* is a connected Riemannian manifold with transitive isometry group.

(2) A metric space X is *locally finite* if for each point $x \in X$, and any $r > 0$, the r-ball around x contains only finitely many points. A graph (X, R) is locally finite if the associated metric is; equivalently $R(a)$ is finite for all $a \in X$. It is *homogeneous* if the isometry group is transitive.

We say that a metric space is (1-)*proper* if each ball (of radius 1) is compact. A 1-proper metric space is automatically complete. A connected Riemannian homogeneous space, or a connected (in the graph sense) locally finite metric space, are proper and hence complete and separable.

Riemannian homogeneous spaces carry a canonical measure (where balls have finite measure). Namely, given $f : U \to \Omega$ where U is an open ball in $\mathbb{R}^n$, we have $\mu(f(U)) = \int_U |J(f)|\, dx$, where $\int_U dx$ is the usual integral, and $J(f)$ is the Jacobian of f, i.e., $\det(\partial_i f / \partial_j x_i)$, computed in any orthonormal basis. Likewise, homogeneous locally finite metric spaces carry a canonical measure: the counting measure, normalized so that a unit ball has measure 1.

Let X be a Riemannian homogeneous space. The isometry group of X is then a Lie group L; and the stabilizer of a point is a compact subgroup of L. See [56, Theorem 1.2].

Conversely, let L be a Lie group and assume given a transitive action of L on a manifold X, with compact point stabilizer K. Then X can be given a Riemannian structure, such that L acts by isometries. (Weyl trick: pick a point p; the stabilizer K of p is compact. Pick any inner product on the tangent space T_pX, and average over K so that we obtain a K-invariant inner product. For any other point q, there exists a unique inner product structure on q such that for any $g \in L$ with $g(p) = q$, g induces an isometry of tangent spaces $T_p X \to T_q X$.) The invariant Riemannian structure is not unique but there is a finite-dimensional space of choices, namely a choice of a K-invariant inner product on the tangent space T_p. The tangent space splits as a direct sum of finitely many invariant subspaces, and the invariant Riemannian structure is determined up to a scalar renormalization on each of them. In any case all these metrics are commensurable, and all commensurable to any L-invariant metric on X.

Like Lie groups, Riemannian homogeneous spaces are rather special creatures: they tend to have no deformations, except for some freedom in constructing nilpotent ones. Two canonical Riemannian homogeneous spaces are associated with each Lie group L: the Lie group itself, and the quotient L/K where K is a maximal compact subgroup (which is unique up to conjugacy). In case the stabilizer of a point acts irreducibly on the tangent space, they were fully classified by Cartan and Wolf.

This leads to examples of homogeneous approximate equivalence relations.

Examples 5.2. (1) Assume (Ω, R) has bounded valency, i.e., $1 \le |R(a)| \le k$. Then $|R^2(a)| \le k^2$ so (Ω, R) is a k'-approximate equivalence relation, $k' \le k^2$.

(2) Assume $|\Omega| \le k^*|R(a)|$ for any a. Then (Ω, R) is a near equivalence relation, and (Ω, R^2) is a k'-approximate equivalence relation. Conversely, a k-approximate equivalence relation *of bounded radius r* is of this type, $|\Omega| \le k^*(|R(a)|)$ with $k^* = k^r$. This is because, inductively, the l-ball $R^l(a) = \bigcup_{c \in R^{l-1}(a)} R(c)$ is a union of $\le k^{l-1}$ 1 balls. When k^* is large compared to k (say $k^* > 10k$), this is still an interesting source of examples for us; for instance if $R(a) = [a-1, a+1]$, it is hard to see locally whether $\Omega = \mathbb{R}$ or $\Omega = \mathbb{R}/k^*$ for some very large k^*.

(3) Let (Ω, d) be a Riemannian homogeneous space. Then balls are not finite but they have finite measure with respect to the canonical measure induced by the Riemannian structure. The metric is k-doubling at scale 1 (and all other scales) for an appropriate k, and all 1-balls have the same measure. We view this as a *model* for k-approximate equivalence relations.

(4) (See Definition 3.36.) Let (Ω, d) be a Riemannian homogeneous space, and $R = \{(x, y) : d(x, y) \le 1\}$. Let (X_i, R_i) be a sequence of symmetric binary relations whose pure probability theory approaches that of (Ω, d). Then using Lemma 4.2, noting that condition (2) of Definition 4.1 holds by approximate homogeneity, we see that (X_i, R_i^2) is an a.e. approximately homogeneous sequence of k'-approximate equivalence relations. We explain below how to obtain such approximations that are *locally finite* (we present this in case $\Omega = L$ has the form $G(\mathbb{R})$, with G a semisimple algebraic group over $\mathbb{R}$).

5.3. ***Sprinkling.*** Let us see how to obtain sequences of locally finite approximations in Examples 5.2(4), by choosing them at random (similar procedures are known as "sprinkling" in some physics literature).

Let L be a Lie group. A *lattice* is a discrete subgroup Λ of L of finite covolume, i.e., L/Λ admits a translation invariant finite measure. Assume that for any compact $K \subset L$ (with $1 \in K$) there exists a lattice Λ_K with $\Lambda_K \cap K = (1)$. This is the case for simple Lie groups L; for $L = G(\mathbb{R})$ an algebraic group over $\mathbb{R}$, where G has no G_m-quotients, one can take the arithmetic lattice $G(\mathbb{Z})$ or a congruence lattice therein (Borel–Harish-Chandra).

Any formula ϕ of local probability logic can quantify only to a bounded distance, i.e., at $x_1, \dots, x_k$ expectation quantifiers are applied only on $\bigcup_i Kx_i$. In this case, if Λ is a sufficiently small lattice (i.e., $KK \cap \Lambda = (1)$), the value of ϕ in L and in L/Λ is the same. Thus it suffices to show how to find finite approximations L/Λ; the pullback to L will be an equally good locally finite approximation to L.

This reduces the problem to the finite volume case. Here a random choice of n points, for large n, provides a good approximation. This is Keisler's relational law of large numbers: in a probability model M (with the measure on M itself, as is the case for L/Λ), the initial sections M_n of a random sequence with the normalized counting measure approach M in the sense that for any sentence ϕ of probability logic, $\phi^M - \phi^{M_n}$ approaches 0. See [52, 6.13; 53, 3.1.3]. In case M has a unique probability logic 1-type, it follows that the sequence M_n is a.e. approximately homogeneous.

Another approach gives a stronger result, at least in the case of an arithmetic lattice, or a cocompact lattice. In the cocompact case, L/Λ along with the metric is easily seen to be interpretable in $\mathbb{R}_{\mathrm{an}}$. In the arithmetic case, [2] prove the existence of an open, semialgebraic *Siegel set* A for L. Within A one can find a fundamental set F for L/Λ, and any element of A can be translated to an element of F via a finite subset of Λ. If C is a compact semialgebraic neighborhood of the identity in L, then CA is contained in finitely many translates of A; this implies that the action of C on L/Λ is also definable semialgebraically. The image in L/Λ of the closed unit ball for the metric may not be semialgebraic, but it is definable in $\mathbb{R}_{\mathrm{an}}$. Thus our structure is interpretable in an o-minimal one, and in particular it is NIP.

Now for a NIP structure M, with a measure μ, a fundamental theorem of Vapnik–Chervonenkis (see [74; 5] and Proposition 2.18) shows that if n points of are chosen at random to give a subset M_n, the law of large numbers holds not just for a given event but uniformly for all definable events. Thus up to ϵ-resolution, the M_n look like M not only to an observer outside both, but also to an internal one at point p who takes into account relative positions to p, or several such points p. The relational version of this follows inductively (much more readily than in [53, 3.1.3], where more careful estimates are needed). It is also easy to see here that when M is homogeneous, the M_n will be approximately homogeneous (not only a.e.).

Remark 5.4. It is natural to ask for a purely probabilistic construction, replacing the use of lattices above. With our present definition of an amenable approximate equivalence relation, sprinkling points on Ω using a Poisson process will not work; while in some sense rare, there would be infinitely many.

Theorem 5.5. *Let* (Ω, R) *be a local ultraproduct of a sequence of approximately homogeneous, amenable, k-approximate equivalence relations. Then there exists a metric space* (X, d_X), *and a surjective* $\alpha : \Omega \to X$, *both canonically defined, such that, letting* $R_X = \{(x, y) \in X^2 : d_X(x, y) \leq 1\}$, *we have*:

(1) *The distance between connected components is bounded strictly above* 0; *each connected component is clopen, and the space* Ξ *of connected components is discrete. The graph induced by* R_X *on* Ξ *is locally finite.*

(2) *Each connected component C of X is a Riemannian homogeneous space*; d_X *is an invariant metric on C.*

(3) *R is commensurable with* $\alpha^* R_X = \{(x, x') : R_X(\alpha(x), \alpha(x'))\}$.

(4) *For* $0 \leq r \leq 1$, *the relation* $d_X(\alpha(x), \alpha(y)) \leq r$ *is* $\bigwedge$*-definable on* Ω. *So is the relation asserting*

$$d_X(\alpha(x), \alpha(y)) \leq r \wedge \alpha(x), \alpha(y) \text{ lie in the same connected component.}$$

If we assume only a.e. approximate homogeneity, the same result holds but the domain of α *is a full measure* $\bigwedge$*-definable set* Ω^*.

A few comments, preliminary to the proof.

(1) The local ultraproduct is taken with respect to the locality relation d_R (see Section 2.3). Thus (Ω, R) is a connected graph by construction. In case only a.e. homogeneity is assumed, we choose a component from Ω^*. Thus the pure probability logic theory does not depend on this choice.

(2) The theorem describes the situation in scales close to 1, and refers to d_X only at such scales. To further study the large-scale structure, one needs to combine the geodesic metric on the Riemannian manifold with the graph metric of R_X.

(3) The proof will begin with the metric ρ of Theorem 4.3, and the completion $\overline{Y}$ with respect to this metric. We would like to find a locally compact group acting on $\overline{Y}$, and connect to the theory of locally compact groups. The proof would be simpler if we assumed full first-order homogeneity, i.e., a unique 1-type, or even a unique 1-type of nonzero measure. Then we could make use of the automorphism group of the saturated model Ω and the induced action on $\overline{Y}$. But this would give a result of quite a different nature, applicable only to finite approximations whose full first-order theory approaches a given limit. Full first-order approximate homogeneity is not really a graph-theoretic condition; it concerns not so much

the given graphs, but all graphs interpretable within them. We prefer therefore to assume only convergence in the sense of probability logic. We must then accept that even though all elements have the same probability logic type, their full types may differ, and the automorphism group of the ultraproduct Ω^* may act trivially on the completion. This obligates us to work with automorphisms of the completion that are not necessarily induced by automorphisms of Ω^*. We will use Theorem 3.30 to obtain automorphisms of a full measure subset Y of $\overline{Y}$.

In the proof of Theorem 5.5, we will pass from the given approximate equivalence relation (Ω, R) to the completion with respect to a metric ρ, and obtain an induced measure. For uniformly continuous functions on the completion, expectation quantifiers can be computed either on the completion or on the original structure, giving the same result. This need no longer be the case for discontinuous functions, such as the characteristic function of R-balls. We thus prepare a smoother version of this function.

Lemma 5.6. *There exists R^* definable from R in pure probability logic, such that*

(1) *R^* is uniformly continuous with respect to the metric ρ.*

(2) *For some $\beta^2 > 0$, if $R(a, b)$ holds, then $R^*(a, b) \geq \beta^2$.*

(3) *If $R^*(a, b) > 0$, then $R^{17}(a, b)$.*

Proof. Let $\alpha_0 : \mathbb{R} \to \mathbb{R}^{\geq 0}$ be the continuous function with value 1 on $\left(-\infty, \frac{1}{4}\right]$, value 0 on $\left[\frac{1}{2}, \infty\right)$, and linear on $\left[\frac{1}{4}, \frac{1}{2}\right]$. Let $\beta > 0$ be a lower bound on the volume of a ρ-ball of radius $\frac{1}{4}$.

Define

$$R^*(x, y) := E_{u,v}\big(\alpha_0(\rho(x, u)) \cdot R^9(u, v) \cdot \alpha_0(\rho(y, u))\big).$$

So R^* is definable from R in pure probability logic.

(1) Uniform continuity is clear since α_0 is uniformly continuous, and ρ is a metric; so if $\rho(x, x') \leq \epsilon$ then $|\rho(x, u) - \rho(x', u)| \leq \epsilon$, and similarly for $\rho(y, v)$.

(2) Assume $R(a, b)$ holds. Whenever $\rho(a, u) \leq 1$ and $\rho(b, v) \leq 1$, we have $R^4(a, u)$ and $R^4(b, v)$ by Theorem 4.3(3), so $R^9(y, v)$. When $\rho(a, u), \rho(b, v) \leq \frac{1}{4}$ we have $\alpha_0(\rho(x, u)) = \alpha_0(\rho(y, v)) = 1$ so the expectation in the definition of R^* is at least the volume of the product of the balls $\rho(a, x) \leq \frac{1}{4}$, $\rho(b, y) \leq \frac{1}{4}$.

(3) If $R^*(a, b) > 0$, then for some u, v we have $R^9(u, v)$ and $\alpha_0(\rho(a, u)) > 0$, $\alpha_0(\rho(b, v)) > 0$, so $\rho(a, u) < \frac{1}{2}$, $\rho(b, v) < \frac{1}{2}$; thus again $R^4(a, u)$ and $R^4(b, v)$ so $R^{17}(a, b)$. $\square$

Proof of Theorem 5.5. By definition of the local ultraproduct, (Ω, R) is connected as a graph. Let ρ be the metric of Theorem 4.3 and let $(\overline{Y}, \bar{\rho})$ be the completion.

By Theorem 4.3(2), $\overline{Y}$ is 1-proper. We have a surjective map $h : \Omega \to \overline{Y}$, such that the pullback of a closed bounded subset of $\overline{Y}^n$ is a bounded $\bigwedge$-definable subset of Ω^n, in pure probability logic, with parameters. Indeed ρ is definable in pure probability logic; if we view ρ as quantifier-free, then for closed $Z \subset \overline{Y}^n$, the pullback $h^{-1}(Z)$ is quantifier-free definable with parameters in (Ω, ρ) alone (see Section 2.5).

Let R^* be as in Lemma 5.6. We consider the structure (Ω, ρ, R^*); it is a reduct (generally a proper reduct) of (Ω, R). For this reduct, ρ can serve as a metric, since both ρ and R^* are uniformly continuous with respect to ρ. By the discussion in Section 2.11, a structure $(\overline{Y}, \bar{\rho}, \overline{R^*})$ is induced; and further we have local expectation quantifiers on this structure, and can speak of the expectation of a pure probability logic definable set (so that a measure on the local type spaces is induced).

Claim 0. *There exists a smallest pure probability logic $\bigwedge$-definable subset Ω^* of Ω of full measure*; *it determines a unique* 1*-type of pure probability logic.*

Proof. The language is countable, and has countably many formulas ϕ in one variable. (We can take ϕ to be in Hoover normal form.) Let $v = v(\phi)$ be the generic value in the sense of Definition 3.36. Then $\phi(x) = v(\phi)$ has full measure by the a.e. almost homogeneity assumption. Let Ω^* be the intersection of $\phi(x) = v(\phi)$ over all ϕ. Then Ω^* has full measure; and the value of any pure probability logic formula is determined on Ω^*. □

Let Y be the image of Ω^* in $\overline{Y}$. Then Y is closed (this can be checked within a given small closed ball, and the topology there is the logic topology), and $\overline{Y} \setminus Y$ has measure 0.

Now ρ is definable from R in local probability logic, so any two elements of Ω^* have the same pure local probability logic type in the language including ρ. It follows that the same is true for their images in Y, in a language including $\bar{\rho}$ and R^*, using the fact that the pullback of $\bar{\rho}$ is ρ, that expectations computed in Y and in $\overline{Y}$ are the same (as $\overline{Y} \setminus Y$ has measure 0), and that the expectation of a function on $\overline{Y}$ equals the expectation of the pullback on Ω, by definition of the measure on Ω (see Section 2.11).

In case the sequence is approximately homogeneous, we have $\Omega = \Omega^*$ and so $Y = \overline{Y}$, since there is only one pure local probability logic type in Ω.

Let $\bar{R}$ be the image of R in $\overline{Y}$, and let R_1 be the image of $R \cap (\Omega^*)^2$ in Y. Form the metrics d_{R_1} on Y and $d_{\bar{R}}$ on $\overline{Y}$. The pullback of a finite $d_{\bar{R}}$-ball $B_r(\bar{a})$ to Ω^* is contained in $B_{5r}(a)$ using Theorem 4.3(3), which controls the thickening caused by taking the ρ-completion and pulling back. Thus all $d_{\bar{R}}$-balls in $\overline{Y}$ have finite measure.

Claim 1. *Y is locally compact*; *all $\bar{R}$-balls have compact closure.*

Proof. Local compactness of $\overline{Y}$ follows from Theorem 4.3(1). Since Y is a closed subset of $\overline{Y}$, it is locally compact too. Now it suffices to show that a $\bar{R}$-ball b is totally bounded with respect to $\bar{\rho}$; i.e., that for any $\epsilon > 0$, b may be covered by finitely many ϵ-balls for $\bar{\rho}$. For this it suffices to see that a maximal set of $\epsilon/2$-$\bar{\rho}$-balls is finite. This in turn follows from the finiteness of $\mu(b)$, and Theorem 4.3(4) (along with the fact that $\mu(\Omega^* \setminus \Omega) = 0$). $\square$

By local compactness the local n-type spaces can be identified with $\overline{Y}^n$ itself. Hence we obtain an induced measure on $\overline{Y}$.

Let $G = \mathrm{Aut}(Y, \bar{\rho}, R^*)$ be the group of isometries of $(Y, \bar{\rho})$ that preserve R^*. G is transitive on Y, by Proposition 3.34. Define a topology on G as in Section 2.11.

Claim 2. *G is locally compact. for $a \in Y$, the point stabilizer $G_a = \{g : ga = a\}$ is compact; for $a \in Y$, the natural map $G/G_a \to Y$ is a homeomorphism.*

Proof. By Lemma 2.15, taking the relation R there to be $\{(x, y) : R^*(a, b) \geq \beta^2\}$, and $d = \rho$. (The assumptions of Lemma 2.15 hold by Claim 1 and Lemma 5.6(2,3).) $\square$

According to Gleason–Yamabe [76], G has an open subgroup H, and a compact normal subgroup N of H (contained in any desired neighborhood of 1), such that H/N is a finite-dimensional Lie group.

Dually to [43], let β_0 be the set of pairs (H, N) with H an open subgroup of G, and N a compact normal subgroup of H. Let β be the set of pairs $(H, N) \in \beta_0$ such that if $(H', N') \in \beta_0$ and $N \leq N' \leq H' \leq H$ then $H = H'$ and $N = N'$. Equivalently, $(H, N) \in \beta$ iff the locally compact group H/N is connected, with no nontrivial compact normal subgroups. (Hence by Yamabe, is a Lie group.) Nonemptiness of β follows from a chain condition on closed subgroups shown in the second paragraph of [43, §4.1]. It is also shown there that if (H, N) and $(H', N') \in \beta$ then $H \cap H'$ has finite index in H and in H', and $(H \cap H', H \cap N') \in \beta$. And H determines N uniquely, i.e., for any H, there is at most one N with $(H, N) \in \beta$.

Fix $a \in Y$. Since G_a is compact while H is open, $G_a \cap H$ has finite index in G_a; in particular there are only finitely many G_a-conjugates of H. Taking their intersection, we see that there exists $(H_a, N_a) \in \beta$ normalized by G_a.

Let $(H_b, N_b) = (gH_ag^{-1}, gN_ag^{-1})$ for any g with $g(a) = b$. Define an equivalence relation E on Y by bEb' iff $H_b = H_{b'}$ and $H_b b = H_{b'} b'$. Then E is G-invariant. If aEa', then $a' = ha$ for some $h \in H_a$. Conversely if $a' = ha$ with $h \in H_a$, then $H_{a'} = hH_ah^{-1} = H_a$; and $H_a a = H_a a' = H_{a'} a'$. Thus the E-class of a is just $H_a a$; it is the image of H_a under the natural map $G \to Y$, $g \mapsto ga$. We saw in Claim 2 that this map induces a homeomorphism $G/G_a \to Y$. Thus the E-class of a is open; by transitivity each E-class is open, and hence each E -class is clopen.

Give Y/E a graph structure via R_1.

Claim 3. *Y/E has finite valency. The topology on Y/E induced from Y is discrete.*

Proof. Since a $d_{\bar{R}}$-ball b is compact while each E-class is open, b meets only finitely many E-classes. □

Since the E-classes are open, there exists $r_0 > 0$ such that the r_0-ball around a is contained in the E-class of a; by transitivity of G, the same holds for all points. Thus two points in distinct E-classes are at ρ-distance $\geq r_0$. We may choose $r_0 \leq \frac{1}{2}$, so that the pullback to Ω of a $\bar{\rho}$ of radius r_0 is contained in finitely many d_R-balls of radius 1 (Theorem 4.3(1)).

Define a finer equivalence relation $\boldsymbol{e}$ on Y by $b\boldsymbol{e}b'$ iff bEb' and $N_b b = N_{b'} b'$. (Note that bEb' implies $H_b = H_{b'}$ and hence $N_b = N_{b'}$.) Then for each E-class $Y' \subset Y$, $Y'/\boldsymbol{e}$ is connected, since a conjugate of the connected group H_a/N_a acts transitively.

Let $X = Y/\boldsymbol{e}$, and $\alpha : \Omega^* \to X$ the composition $\Omega^* \to Y \to X$. Define

$$d_X(x, y) := \inf\{\min(r_0, \bar{\rho}(a, b)) : a, b \in Y, \alpha(a) = x, \alpha(b) = y\}.$$

Since the classes of $\boldsymbol{e}$ are compact, the infimum in this definition is attained, and thus $d_X(x, y) = 0$ implies $x = y$. It is clear that $d_X(x, x) = 0$ and that d_X is symmetric. Let us consider the triangle inequality $d(x, z) \leq d(x, y) + d(y, z)$. If $d(x, y) \geq r_0$ or $d(y, z) \geq r_0$ the inequality is clear; so we may assume $d(x, y), d(y, z) < r_0$. Let (a, b) attain the minimum in the definition of $d_X(x, y)$, and likewise b', c for y, z. Then $\bar{\rho}(a, b) < r_0$ so aEb, and likewise $b'Ec$. We also have $b\boldsymbol{e}b'$ since $\alpha(b) = y = \alpha(b')$. Thus $b' = nb$ for some $n \in N_b = N_a = N_c$. So $\bar{\rho}(b', c) = \bar{\rho}(b, c')$, where $nc' = c$. It follows that

$$d_X(x, z) \leq \bar{\rho}(a, c') \leq \bar{\rho}(a, b) + \bar{\rho}(b, c') = d_X(x, y) + d_X(y, z).$$

Hence d_X is a metric on X. It induces the same topology on X as the quotient topology inherited from Y via $X = Y/\boldsymbol{e}$.

The pullback of an $r_0/2$-ball a of X to Y is the union over a $\boldsymbol{e}$-class b of the $r_0/2$-ρ-balls around elements of b. Since b is compact, finitely many of these $r_0/2$-balls cover b, so that the union is contained in a finite union of r_0-ρ-balls. Hence R is commensurable with $\alpha^* R_X$, where $R_X = \{(x, x') \in X^2 : d_X(x, x') \leq r_0/2\}$. As all the clauses of Theorem 5.5 except for (3) are invariant under rescaling, we may replace d_X by $2d_X/r_0$, so as to obtain (3).

The connected components of X are the images of the classes of E, and are at distance at least r_0 from each other. On each connected component, we have a transitive action of a Lie group $\mathsf{L} \cong H_a/N_a$, respecting the metric. The induced metric on the space of connected components is locally finite by Claim 3.

Let us now address the definability issues (4). Recall the map $h : \Omega \to \overline{Y}$ from the first lines of the proof. The equivalence relation $h(x) = h(y)$ is $\bigwedge$-definable, by the pure probability formula $\rho(x, y) = 0$, and more generally $\bar{\rho}(hx, hy) = \rho(x, y)$. Thus $(\overline{Y}, Y, \bar{\rho})$ can be viewed as interpretable in (Ω, R).

Let $P \subset Y^k$ be an automorphism-invariant closed relation on Y. Pick $a \in Y$, and consider a bounded $\bar{\rho}$-ball B around a, say of radius $r_1 \leq 1$. Then B is compact; so the restriction $P|B = P \cap B^k$ is a-definable; say via a formula $\psi_P(x, a)$ of continuous logic. Since G acts transitively on Y, $\psi_P(x, y)$ defines P on Y, provided P is local, i.e., $P(x_1, \dots, x_k)$ implies $\rho(x_i, x_j) \leq r_1$. In particular this is the case for $P = \boldsymbol{e}$, since the classes of $\boldsymbol{e}$ are compact, and for the relations in the statement of (4). Pulling back to Ω^* we see that P is $\bigwedge$-definable (though not necessarily by a formula of pure probability logic). $\square$

Example 5.7. Let A_n be the interval $[-n, n]$, with Lebesgue measure, and let $R(x, y)$ be the relation $|x - y| \leq 1$. This sequence of 2-approximate equivalence relations is not strictly approximately homogeneous; the measure of $\{x : R(a, x)\}$ is 2 towards the middle, but approaches 1 near the endpoints. However, even in full continuous logic with probability quantifiers, it is a.e. approximately homogeneous; a formula whose quantifiers look out to distance d will take the same value on all points except for the intervals $[-n, d - n]$ and $[n - d, n]$, whose measure $2k/n$ approaches 0 with n.

5.8. ***Definability and asymptotic structure.*** The definability statements (4) in Theorem 5.5 are crucial for deducing asymptotic consequences from the structure of the limit. We plan to return to this in the future, and for now only sketch a basic statement by way of illustration. Let (A_n, R_n) be a sequence of increasingly homogeneous locally finite graphs. Fix $r > 0$ and let B_n be a ball of radius r in A_n, and B_∞ a ball of radius r in an ultraproduct A. By assumption the B_n look increasingly similar as $n \to \infty$.

Let E be the formula defining the "same connected component" equivalence relation on B_∞. Then E has a finite number f of classes on B_∞. For almost all n, the same formula E defines a partition of B_n into f classes. One can further find a modified graph structure R_n' on B_n, uniformly commensurable to R_n. Let C_n be one of the f classes of B_n. After refinement of the indices n, (C_n, R_n') converge to a bounded region on a Riemannian homogeneous space H_r, with graph structure "distance ≤ 1" relative to the Riemannian metric d_{Riemann}. One can also find asymptotic versions of the metric d_{Riemann} on the C_n.

5.9. ***Homogeneity.*** Consider an approximate equivalence relation (Ω, R) with transitive automorphism group G. Let H be the stabilizer of a point $a \in \Omega$. Let $X = \{g \in G : (a, ga) \in R\}$. It is easy to see that $X = X^{-1}$ is an approximate subgroup of G, and $HX = X$.

Conversely, a triple (G, X, H) with X an approximate subgroup of G, H a subgroup with $HX = X$ gives an approximate equivalence relation $\mathcal{R}_{G,X,H}$ on G/H with transitive automorphism group. Namely, $(g_1 H, g_2 H) \in \mathcal{R}_{G,X,H}$ iff $g_1^{-1} g_2 \in X$. When $H = 1$ we write $\mathcal{R}_{G,X}$ or just $\mathcal{R}_X$.

Problem 5.10 (rigidity). Let G be a semisimple Lie group G, with a maximal compact subgroup K. Let $R(x, y)$ be the relation: $y^{-1}x \in K$. Show that the pure probability logic theory of (G, R) determines G uniquely. Further, does there exist a single sentence σ of probability logic such that any increasingly homogeneous sequence of approximate equivalence relations X_i whose $\sigma(X_i)$ converges to $\sigma(G)$ must in fact converge to G? This would say that a resident of X_i (for large enough i) can reasonably guess the limit G that the sequence is tending to.

Assume the ultraproduct M of Theorem 5.5 has an associated Riemannian symmetric space whose Lie group L is simple, with a finitely presented lattice Λ, with generators $\lambda_1, \dots, \lambda_k$. Describe the element λ_i of L as a limit of "rough" symmetries of the approximating graphs, and how to recognize the relations on the λ_i from rough relations among these. When this can be done and Λ is simple, it should becomes possible to recognize it by looking at a single sufficiently good approximation.

Problem 5.11. Theorem 5.5 describes completely the connected case, and the locally finite case. But the mixed case is not fully described. What are the possible homogeneous extensions (Ω, R) of a homogeneous Riemannian space X by a homogeneous, locally finite graph Ξ?

We have copies X_a of X, for $a \in \Xi$. Say $G = \mathrm{Aut}(X)$ is a centerless semisimple Lie group. In this case, for two points a, b of Ξ connected by an edge, there exists a unique isometry $X_a \to X_b$ at finite d_R-distance. This gives a homomorphism $\pi_1(X, a) \to \mathrm{Aut}(X_a)$. Does this fully describe the structure of (Ω, R), up to commensurability?

Problem 5.12. Theorem 5.16 of [41] describes the structure of approximate subgroups without an amenability assumption. Generalize this and Theorem 5.5 to homogeneous approximate equivalence relations; see Section 5.9. Definability may be challenging as the definability statement in [41, Theorem 5.16] is weaker than in the amenable case; and it refers a priori to the group G, a second-order object from the point of view of (Ω, R).

Problem 5.13. In the model-theoretic limit, the pieces of a (well-chosen) partition of any structure into n pieces tends to a union of *homogeneous* structures. Given an arbitrary approximate equivalence relation, can one partition into pieces that resemble various Riemannian homogeneous spaces? Here is a more precise formulation:

Let $(M_i)_{i \in \mathbb{N}}$ be any family of locally finite, k-approximate equivalence relations; we make no homogeneity assumption. Let M be a (sufficiently saturated) ultraproduct; so $\mathrm{Aut}(M)$ acts transitively on each complete type P. The restriction to P is then a homogeneous k-approximate equivalence relation, and a structure theorem should apply (either using an answer to Problem 5.12, or disintegrating the measure over almost all P). What does this imply for the given M_i?

One might expect a statement of the following type: take any function $\alpha : \mathbb{N} \to \mathbb{N}$ growing to ∞, as slowly as you like. Then one can find a subsequence I of $\mathbb{N}$, and for $i \in I$ a partition of M_i into $\alpha(i)$ definable sets $(D_{i,k} : k < \alpha(i))$, refining the partitions $D_{i',k'}$ for $i' < i$; such that each branch resembles, more and more closely, a k-approximate equivalence relation commensurable to a Riemannian homogeneous space. By a branch we mean a choice of a definable set $D_{i,k(i)}$ for each i, so that $D_{i,k(i)}$ implies $D_{i',k(i')}$ if $i' < i \in I$; it gives a sequence of graphs $D_{i,k(i)}(M_i)$.

See [10] for the 1-dimensional case.

6. Strong approximation: from groups to graphs

This section again addresses the structure of approximate equivalence relations; but here we assume definability over finite fields, or more generally the existence of a dimension theory similar to the one available for pseudofinite fields. We will see that approximate equivalence relations in this setting are close to actual equivalence relations.

We first recall a key statement of [44] in the case of groups, Theorem 6.1 below, that forms the model of the graph-theoretic generalization we aim for. This was the main ingredient in model-theoretic proofs of *strong approximation* over prime fields, for instance of the fact that if H is a Zariski dense subgroup of $\mathrm{SL}_n(\mathbb{Z})$ then H maps surjectively to almost every $\mathrm{SL}_n(\mathbb{F}_p)$.

Theorem 6.1 (strong approximation lemma: groups). *Let $F = \mathbb{F}_p$, p nonstandard. Let G be a definable group over F, and let $(X_i : i \in I)$ be a family[6] of definable subsets of G.*

(A) *There exists a definable H such that*

  (1) *H is a subgroup of G.*
  (2) *$H \subset \langle \bigcup X_i \rangle$.*
  (3) *X_i/H is finite.*

  If read for standard primes p, each X_i should be definable uniformly in p; and in (3), *each X_i/H has finite cardinality independent of p. Moreover, if G is an algebraic group*:

(B) *There exists a homomorphism $h : \widetilde{H} \to G$ of algebraic groups, with finite kernel such that $H = h(\widetilde{H}(F))$.*

6.2. Here is how strong approximation follows from Theorem 6.1. Let G be a linear algebraic group. Applied to the family of one-dimensional unipotent subgroups X_i

[6] Not necessarily uniformly definable.

of *an arbitrary subgroup* Γ *of* G, Theorem 6.1 shows that Γ contains a *definable* normal subgroup H (generated by the unipotent elements of Γ) such that Γ/H has no unipotent elements. By Jordan's theorem it follows that Γ/H is abelian-by-bounded. In particular, the image of a Zariski dense subgroup of $\mathrm{SL}_n(\mathbb{Z})$ in $\mathrm{SL}_n(\mathbb{F}_p)$ has bounded index, and admits the description of in Theorem 6.1(B). See [44, Proposition 4.3 and 7.3]. These results were previously proved by other means by Weisfeiler, Nori, and Gabber.

6.3. ***Generalization to graphs.*** The graph version applies to definable graphs over F. For instance, the graph may be (V, X) with V a variety and X a subvariety of V^2, or more generally the projection of the rational points of a variety W mapping onto V^2. We try to study them *up to graphs of bounded valency*. Essentially we show that after a partition of the vertices into a (bounded) finite number of pieces, and after fibering each piece over a graph of bounded valency in each piece, X generates an equivalence relation on each piece in a bounded number of steps. Between any two distinct pieces, the induced graph has finite valency in at least one direction, following an interesting partial ordering of the set of pieces.

Although Theorem 6.6(B) is formulated for pseudofinite fields, part (A) and Proposition 6.4 are valid for structures with a finite, definable S1-rank in the sense of [40]. We refer to such structures and their theories as S1-structures and theories.

The reader who wishes can read both parts for pseudofinite fields.

For a graph (G, X), $X \subset G^2$, we let $\sim_G$ be the equivalence relation of connectedness, i.e., $\sim_G$ is the smallest equivalence relation on G containing X.

Proposition 6.4. *Let* (G, X) *be a definable graph in an* $S1$*-theory. On each type* P *of* G *there exists a canonical* $\bigwedge$*-definable equivalence relation* E_P *such that* $E_P \subset \sim_G$, *in fact* $E_P \subset X[m]$ *for some* m, *and such that the induced graph on* P/E_P *is locally finite.*

The definition of E_P *depends only on* P *and on* $\sim_G$.

Remark 6.5. Proposition 6.4 is valid more generally if $X = \bigcup_i X_i$ is only $\bigvee$-definable; i.e., the edges X are given as a countable union of definable sets in a countably saturated model. In this case local finiteness means that any finite union $X' = \bigcup_{i \in F} X_i$, and any $\bar{a} \in P/E_P$, there are only finitely many $\bar{b} \in P/E_P$ with $(\bar{a}, \bar{b}) \in X'/E_P$.

Here is a more detailed version, that allows passage to the finite.

Theorem 6.6 (strong approximation lemma: graphs).

(A) *Let* (G, X) *be an* $S1$*-structure with* $X \subset G^2$ *symmetric, reflexive. Then there exists a definable partition* $G = D_1 \cup \cdots \cup D_n$ *a definable function* f *on* G, *and* $m \in \mathbb{N}$ *such that*:

(1) *If* $f(a) = f(b)$ *then* $d_X(a, b) \leq m$.

(2) *For* $a \in G$, $f(X(a))$ *is finite.*

(3) *For* $i \leq j \leq n$, *let* $\overline{X}_{i,j} = \{(f_i(x), f_j(y)) : x \in D_i, y \in D_j, (x, y) \in X\}$. *Then for any* $\bar{b} \in f(D_j)$ *there are only finitely many* $\bar{a} \in f(D_i)$ *with* $(\bar{a}, \bar{b}) \in \overline{X}_{ij}$.

(B) *Let* F *be an ultraproduct of finite fields. Let* (G, X) *be an* F*-definable graph*: G *is a definable subset of* $V(F)$, V *an* F*-variety, and* $X \subset G^2$ *is definable. Then (A) holds with an* ***algebraic*** f *and* D_i: *we can find a quasifinite morphism of* F*-varieties* $\rho : \widetilde{G} \to V$ *with* $\rho(\widetilde{G}(F)) = G(F)$, *and regular functions* ϕ *on* $\widetilde{G}$, *such that for* $u \in \widetilde{G}$,

$$f(\rho(u)) = \phi(u).$$

Preliminaries to proof. We work in a sufficiently saturated and homogeneous model. Two elements with the same type are then conjugate by some automorphism. We write $d(b/a)$ for the dimension of the smallest a-definable Zariski closed set including b. More generally in the setting of finite S1-rank, $d(b/a)$ is the least S1-rank of a formula $\phi(x, a)$ true of b. We write $a \mathop{\smile\!\!\!\!\!\!\bigcup}_c b$ if $d(a/b, c) = d(a/c)$. In the case of pseudofinite fields, dependence and independence are determined by the quantifier-free type; we have $a \mathop{\smile\!\!\!\!\!\!\bigcup}_c b$ iff the algebraic locus of $a/\operatorname{acl}(b, c)$ (namely the variety V of smallest dimension defined over $\operatorname{acl}(b, c)$ with $a \in V$) is already defined over b.

Canonical bases. Given tuples b, c there exists a unique smallest algebraically closed $\bar{c} \subset \operatorname{acl}(c)$ with $b \mathop{\smile\!\!\!\!\!\!\bigcup}_{\bar{c}} c$. We write $\bar{c} = \operatorname{CB}(b/c)$. If $\operatorname{acl}(c) \subset \operatorname{acl}(c')$ and $b \mathop{\smile\!\!\!\!\!\!\bigcup}_c c'$ then $\operatorname{CB}(b/c') = \operatorname{CB}(b/c)$. In the case of pseudofinite fields, $\operatorname{CB}(b/c)$ is the field of definition of the locus of b over c.

In simple theories of finite SU-rank, one can instead use canonical bases in the sense of simple theories [37].

As a consequence of the existence of canonical bases, we see that if $b \mathop{\smile\!\!\!\!\!\!\bigcup}_d c$ and $b \mathop{\smile\!\!\!\!\!\!\bigcup}_{d'} c$, where $d, d' \subset \operatorname{acl}(c)$, and $\operatorname{acl}(d'') = \operatorname{acl}(d) \cap \operatorname{acl}(d')$, then $b \mathop{\smile\!\!\!\!\!\!\bigcup}_{d''} c$.

Elimination of imaginaries. $\mathbb{F}$ has a unique extension $\mathbb{F}_n$ of degree n; it is Galois with Galois group $\mathbb{Z}/n\mathbb{Z}$, and when F is finite, $\operatorname{Aut}(F_n/F)$ has a canonical element ϕ_n. We view this as part of the structure of each finite field F. Correspondingly, we obtain an element ϕ_n of $\operatorname{Aut}(\mathbb{F}_n/\mathbb{F})$, that we view as definable. This amounts to naming a certain algebraic imaginary element e_n coding ϕ_n, for each n, on top of the field structure. With this understood, $\operatorname{Th}(\mathbb{F})$ admits elimination of imaginaries; see [22].

Proof of Proposition 6.4. For $u, v \in G$, write $u \sim v$ if u, v lie in the same connected component of the graph (G, X). This is a countable union of definable relations,

$$\bigvee_{j \in \mathbb{N}} d_X(u, v) \leq j.$$

Fix for a moment an element $b \in G$, with type P. Let $W = W_b$ be the connected component of b within the graph (G, X). Choose $a \in W$ with $d(b/a)$ as large as possible; let $m_P = d(b/a)$. Let $\delta_P = d_X(a, b)$ (for definiteness, choose a so that δ_P is minimal, subject to $d(b/a) = m_P$). Let $e = \mathrm{CB}(b/a)$.

Claim 1. *For any* $c \in W$, $e \in \mathrm{acl}(c)$. *In particular*, $e \in \mathrm{acl}(b)$.

We first prove the claim under the assumption that $c \mathop{\smile\!\!\!\!\!\!\!\;\lower0.5ex\hbox{$\mid$}}_a b$. We have $d(b/c) \geq d(b/ac) = d(b/a)$. By maximality of $d(b/a)$, $d(b/c) = d(b/a) = d(b/ac)$. So $b \mathop{\smile\!\!\!\!\!\!\!\;\lower0.5ex\hbox{$\mid$}}_c a$. Hence $e = \mathrm{CB}(b/a) = \mathrm{CB}(b/ac) \in \mathrm{acl}(c)$.

Now for a general c, take c' with $\mathrm{tp}(c/\mathrm{acl}(a)) = \mathrm{tp}(c'/\mathrm{acl}(a))$ and $c' \mathop{\smile\!\!\!\!\!\!\!\;\lower0.5ex\hbox{$\mid$}}_a b$. Since $a \sim c$ we have also $a \sim c'$, so $c' \in W$. By the special case just proved, we have $e \in \mathrm{acl}(c')$. Since $e \in \mathrm{acl}(a)$, we have $\mathrm{tp}(c/e) = \mathrm{tp}(c'/e)$. Thus $e \in \mathrm{acl}(c)$.

Claim 2. *If* $\mathrm{tp}(b'/\mathrm{acl}(e)) = \mathrm{tp}(b/\mathrm{acl}(e))$ *and* $b \mathop{\smile\!\!\!\!\!\!\!\;\lower0.5ex\hbox{$\mid$}}_e b'$ *then* $b' \in W$; *in fact*, $d_X(b, b') \leq 2\delta_P$.

Recall $e = \mathrm{CB}(b/a)$ so $b \mathop{\smile\!\!\!\!\!\!\!\;\lower0.5ex\hbox{$\mid$}}_e a$. Using the independence theorem over $\mathrm{acl}(e)$ (computed in M^{eq}), we can find a' with $\mathrm{tp}(a', b/\mathrm{acl}(e)) = \mathrm{tp}(a', b'/\mathrm{acl}(e)) = \mathrm{tp}(a, b/\mathrm{acl}(e))$. By the definition of δ_P and the choice of a we have $d_X(b', a) = d_X(b, a) = \delta_P$. Since $\mathrm{tp}(a', b) = \mathrm{tp}(a', b') = \mathrm{tp}(a, b)$ we also have $d_X(a', b') = d_X(a', b) = d_X(a, b) = \delta_P$. So $d_X(b, b') \leq 2\delta_P$.

Claim 3. *If* $\mathrm{tp}(b'/\mathrm{acl}(e)) = \mathrm{tp}(b/\mathrm{acl}(e))$ *then* $b' \in W$; *in fact*, $d_X(b, b') \leq 4\delta_P$.

Indeed let $\mathrm{tp}(b''/\mathrm{acl}(e)) = \mathrm{tp}(b/\mathrm{acl}(e))$ with $b'' \mathop{\smile\!\!\!\!\!\!\!\;\lower0.5ex\hbox{$\mid$}}_e b, b'$. By Claim 2 we have $d_X(b, b'') \leq 2\delta_P$ and $d_X(b', b'') \leq 2\delta_P$; so $d_X(b, b') \leq 4\delta_P$.

Claim 4. $\mathrm{acl}(e) = \bigcap_{c \in W} \mathrm{acl}(c)$.

We already saw one direction in Claim 1. Conversely in Claim 2 we saw that if $\mathrm{tp}(b'/\mathrm{acl}(e)) = \mathrm{tp}(b/\mathrm{acl}(e))$ and $b \mathop{\smile\!\!\!\!\!\!\!\;\lower0.5ex\hbox{$\mid$}}_e b'$ then $b' \in W$; if $d \in \bigcap_{c \in W} \mathrm{acl}(c)$, then $d \in \mathrm{acl}(b) \cap \mathrm{acl}(b')$ so $d \in \mathrm{acl}(e)$.

Claim 5. *Let* $\mathrm{tp}(e'/b) = \mathrm{tp}(e/b)$; *then* $\mathrm{acl}(e') = \mathrm{acl}(e)$. *More generally if* $\mathrm{Aut}(M/b')$ *leaves* W *invariant, and* $\mathrm{tp}(e'/b') = \mathrm{tp}(e/b')$, *then* $\mathrm{acl}(e') = \mathrm{acl}(e)$.

Indeed let σ be an automorphism fixing b' and with $\sigma(e) = e'$; then $\sigma(W) = W$, and by Claim 4 we have

$$\mathrm{acl}(e') = \sigma(\mathrm{acl}(e)) = \bigcap_{c \in \sigma(W)} \mathrm{acl}(c) = \bigcap_{c \in W} \mathrm{acl}(c) = \mathrm{acl}(e).$$

Claim 6. *There exists a definable function* f_P *such that* $\mathrm{acl}(e) = \mathrm{acl}(f_P(b))$.

Let $\boldsymbol{e}$ be a code for the finite set $\bar{E}$ of conjugates of e/b. Then $\boldsymbol{e} \in \mathrm{dcl}(b)$. So $\boldsymbol{e} = f_P(b)$ for some definable function f_P. By Claim 5, each element of $\bar{E}$ is equialgebraic with e; so we have also $\mathrm{acl}(e) = \mathrm{acl}(\boldsymbol{e})$.

Thus far, the element e played a role only via $\operatorname{acl}(e)$; so we may replace e by $\boldsymbol{e}$ with no loss.

Define an equivalence relation E_P on P, setting $x E_P y$ iff $f_P(x) = f_P(y)$ and $\operatorname{tp}(x/\operatorname{acl}(f_P(x))) = \operatorname{tp}(y/\operatorname{acl}(f_P(x)))$. It is easy to see that E_P is $\bigwedge$-definable. Both equations together can be written as

$$\operatorname{tp}(x/\operatorname{acl}(f_P(x))) = \operatorname{tp}(y/\operatorname{acl}(f_P(x))).$$

(Since this implies $f_P(x) = f_P(y)$.) As this equation refers only to $\operatorname{acl}(f_P(x)) = \bigcap_{u\sim x} \operatorname{acl}(u)$, it does not depend on the specific choice of the 0-definable function f_P.

Claim 7. *If $c, d \in P$ and $c \sim d$ then $f_P(c) \in \operatorname{acl}(f_P(d))$.*

Proof. Let d' be independent from c, d over $f_P(c)$, with $\operatorname{tp}(d'/\operatorname{acl}(f_P(d))) = \operatorname{tp}(d/\operatorname{acl}(f_P(d)))$. In particular $d' \in P$, $f_P(d') = f_P(d)$, and $d' E_P d$. By Claim 3, $d' \in W_d$. So $c \sim d'$. By Claim 1, $f_P(c) \in \operatorname{acl}(d')$. Since $c \mathrel{\raise0.2ex\hbox{$\bigcup$}\!\!\!\!\!\lower0.2ex\hbox{$-$}}_{f_P(d)} d'$, we have $f_P(c) \in \operatorname{acl}(f_P(d))$. □

Local finiteness follows from Claim 7 and compactness. In particular each $\sim$-class on P is a countable union of E_P-classes.

This ends the proof of Proposition 6.4. □

Proof of Theorem 6.6. We continue with the proof of Theorem 6.6.

The equivalence relation E_P is an intersection of 0-definable equivalence relations; so for one of these equivalence relations E_0, for some 0-definable set D with $b \in D$, we have that for all $b', b'' \in D$, with $\delta = \delta_P$,

$$b' E_0 b'' \text{ implies } d_X(b', b'') \le 4\delta. \tag{$*$}$$

Let $f(b) = b/E_0$. A conjugate of $f(b)$ has the form $f(b')$; a conjugate over $\operatorname{acl}(f_P(b))$ has the form $f(b')$ where $\operatorname{tp}(b/\operatorname{acl}(f_P(b))) = \operatorname{tp}(b'/\operatorname{acl}(f_P(b)))$; and so since E_0 refines this equivalence relation, we have $b E_0 b'$, so $f(b) = f(b')$. Thus $f(b) \in \operatorname{acl}(f_P(b))$.

On the other hand, we saw that if c, b are connected then $e \in \operatorname{acl}(c)$ and so $\boldsymbol{e} \in \operatorname{acl}(c)$. In particular, if $d_X(b, c) \le 5\delta_P$ then $\boldsymbol{e} \in \operatorname{acl}(c)$. Thus we can also choose D so that for all $b' \in D$, $c' \in G$, with $\delta = \delta_P$ we have

$$\text{if } d_X(b', c') \le 5\delta \text{ then } f(b') \in \operatorname{acl}(c') \tag{$**$}$$

(here, $f(b') \in \operatorname{acl}(c')$ can be replaced by a single formula, obtainable by compactness, that guarantees it).

Let F be the set of all triples (D, f, δ) with the above two properties. We saw that any b lies in D for some $(D, f, \delta) \in F$. By compactness there exists finitely many such pairs (D_i, f_i, δ_i) such that $\bigcup_i D_i = G$. We may refine them so as to be disjoint, and then take the union to obtain a single function f defined on G, into the disjoint union of the ranges of f_i; so a fiber of f is a fiber of some f_i on D_i, and thus (∗) holds with $\delta = \delta_i$ and (∗∗) holds with $D = G$, $\delta = \delta_i$.

$$\text{If } b, b' \in D_i,\ f(b) = f(b_i),\ \text{ then } d_X(b', b'') \leq 4\delta_i. \tag{$*_i$}$$

$$\text{If } b' \in D_i, c' \in G, d_X(b', c') \leq 5\delta_i,\ \text{ then } f(b') \in \operatorname{acl}(c'). \tag{$**_i$}$$

Define $H = \{(x, y) \in G : f(x) = f(y)\}$. By $(*_i)$, since $f(x) = f(y)$ implies x, y lie in the same D_i, (1) holds with $m = \max \delta_i$. Similarly (2) follows from the $(**_i)$ and compactness (using $1 \leq \delta_i$ for each i).

Towards (3), reorder the D_i so that $\delta_i \geq \delta_j$ for $i \leq j$. Fix $i \leq j$, and let $b' \in D_i$, $c' \in D_j$. If $d_X(b', c') \leq 1$ then $f(b') \in \operatorname{acl}(c'')$ for any $c'' \in D_j$ with $f(c'') = f(c')$; the reason is that $d_X(c', c'') \leq 4\delta_j$ by $(*)_j$, so $d_X(b', c'') \leq 4\delta_j + 1$ and by $(**_i)$, since $4\delta_j + 1 \leq 5\delta_i$, we have $f(b') \in \operatorname{acl}(c'')$. Thus $f(b') \in \operatorname{acl}(f(c'))$. Item (3) follows.

Let us now prove part (B). We work over a subfield F_0 so that F has elimination of imaginaries, and model completeness in the form: every definable set is a projection of the set of rational points of a variety under a finite morphism. See Remarks 6.7(8) for a better statement with control of F_0.

Thus the equivalence relation $(x, y) \in H$ can be written as $f(x) = f(y)$ for some definable function f into F^k. For each complete type P of G, one can find a variety G_P, a morphism $h'_P : G_P \to \mathbb{A}^k$ with finite fibers, and regular functions ϕ'_P on G_P such that for any $a \in P$, for some b we have $h'_P(b) = a$ and $f(a) = \phi'_P(b)$. This is a form of the algebraic boundedness of pseudofinite fields ([31]). It follows from Ax's theorem applied to the graph of f. (This is a finite projection of the rational points of a quantifier-free type of ACF, which can be enlarged to be a locally closed subset of some variety; but this is then itself a variety.)

By Ax's theorem, any definable subset of G_P has the form $g(\widetilde{G}_P)$ for some variety $\widetilde{G}_P$ and morphism g with finite fibers. In particular this is true for $\{y \in G_P : \phi'_P(y) = f(h'_P(y))\}$. Define ϕ_P on $\widetilde{G}_P$ by $\phi_P = \phi'_P \circ g$, and $\rho_P = h'_P \circ g$. Then $f(\rho_P(x))$ is given by regular functions, and $P \subseteq \rho_P(\widetilde{G}_P)$. By compactness, we can find a finite number of $\rho_i : \widetilde{G}_i \to \mathbb{A}^k$ such that $f(\rho_i(x))$ is given by regular functions on each $\widetilde{G}_i$, and $\bigcup_i \rho_i(\widetilde{G}_i) = G$. Let $(\widetilde{G}, \rho)$ be the disjoint union of $(\widetilde{G}_i, \rho_i)$. Then $\rho : \widetilde{G} \to \mathbb{A}^k$ has finite fibers and image G, and there exists a tuple of regular functions ϕ on $\widetilde{G}$, such that $\phi(x) = f(\rho(x))$ for $x \in \widetilde{G}$; the "moreover" follows. □

Remarks 6.7. (1) The "moreover" is made explicit as follows: say $G \subset F^k$. Then there exists a morphism of varieties $h : \widetilde{G} \to \mathbb{A}^k$ with finite fibers, and a tuple of regular functions ϕ on $\widetilde{G}$, such that $\phi(x) = \phi(x')$ if $h(x) = h(x')$, and

$$H = \{(h(x), h(y)) : x, y \in \widetilde{G}(F), \phi(x) = \phi(y)\}.$$

In particular, each H-class has the form $h(\phi^{-1}(c) \cap \widetilde{G}(F))$ for some c.

(2) One can add that piecewise on G (i.e., on each of finitely many definable pieces G_ν), the rational invariants suffice to determine a class of H up to finitely many possibilities. In other words there are rational functions ψ on G_ν, such that the equivalence relation H' defined by $\psi(x) = \psi(y)$ contains $H|G_\nu$, and each H'-class is a finite (bounded) union of H-classes. The $\psi(x)$ will list the coefficients of the minimal polynomials satisfied by the $\phi(y)$ for $h(y) = x$.

(3) The theorem can be stated for the family of finite fields, in place of a single pseudofinite field. Then one adds uniformity to the definability assumption and to the conclusion. In particular m is bounded, that the valencies in (2) are uniformly bounded, and the complexity of h, $\widetilde{G}$, ϕ is bounded independently of p.

(4) Note in particular that for each $i \le n$, $X \cap D_i^2$ induces a finite valency graph on $f(D_i)$. Globally on $f(G)$, there is a multilayered structure with layers $f(D_i)$; the graph has finite valency in the downwards direction, regarding arrows going from $f(D_j)$ to $f(D_{\le j})$, but not necessarily upwards. This kind of tree-like structure is intrinsic and cannot be avoided; see Example 6.13.

(5) Part (A) of Theorem 6.1 admits an analogue for stable or even simple theories, near a given regular type p, using semiregular p-weight in place of dimension. Is there a similar generalization of Theorem 6.6? (It seems likely.)

(6) For simple Robinson theories, even of SU-rank one, it was shown in [39] that E/E_P may be a connected compact space. But pseudofiniteness was not considered there, and it would be interesting to see what happens when this hypothesis is added.

(7) Part (B) of Theorem 6.6 is valid for structures interpretable in PF, not necessarily with the full induced structure. (Thus the transitivity assumption is easy to attain, given a group action.)

(8) In part (B), we needed to adjoin constants so as to have model completeness as in Ax. If in the statement we replace $V(F)$ by $V(F_n)$, where F is a (periodic) difference variety and F_n is the degree n extension of F, or alternatively Artin symbols, this becomes unnecessary; see [40]. In this formulation the theorem is valid over F_0^{alg} where F_0 is the field of 0-definable elements of F. The reason for going to F_0^{alg} rather than F_0 is the need to name the algebraic imaginary parameters ϕ_n, coding a generating automorphism of the Galois group of the degree n extension of F, in order to have elimination of imaginaries; these are definable over F_0^{alg}.

(9) Part (A) of Theorem 6.1 can easily be recovered from part (A) of Theorem 6.6. Part (B) of Theorem 6.6 partially generalizes part (B) of Theorem 6.1, but to obtain a group structure on the covering variety appears to require additional argument.

6.8. ***Definability of connected components.*** We showed in general that connectedness over $\mathbb{F}_p$ of a definable graph can be definably reduced to a locally finite graph. Here we impose stronger assumptions that allow ruling out the locally finite graph, and thus showing definability of the connected components. This will be the graph analogue of Section 6.2. In particular under these assumptions, if the graph generated by X on $\mathbb{F}_p^n$ fails to be connected for infinitely many p, then this is so for an algebraic reason.

Recall that a definable group is *connected* if it has no proper definable subgroups of finite index. If F is a pseudofinite field and G is a simply connected algebraic group, then $G(F)$ is connected ([44]).

Let T be an S1-theory. Let $\mathcal{G}$ be an ind-definable group with a transitive action on a definable set (V, X). Let I be an index set, and for $i \in I$, let $(J_i(v) : v \in V)$ be a definable family of definable subgroups of $\mathcal{G}$. Assume conjugation invariance: $J_i(gx) = gJ_i(x)g^{-1}$. Let $\sim_V (a)$ denote the connected component of $a \in V$.

Corollary 6.9. *Assume that in any model of T, whenever $a \in V$, $J_i(a)$ is a connected definable group, $J_i(a) \cdot a \subseteq \sim_V (a)$, and if $(a, b) \in X$ then $j \cdot a = b$ for some $i \in I$ and some $j \in J_i(a)$, or dually. Then $\sim_V$ is definable.*

Proof. By assumption, $\mathcal{G}$ acts on (V, X) by automorphisms; so as a graph, (V, X) has a unique type P. Let $E = E_P$ be the equivalence relation of Theorem 6.6. Then the $\mathcal{G}$-action preserves E; and V/E is a locally finite graph. Now for $a \in V$, and $i \in I$, by assumption $J_i(a) \cdot a \subseteq \sim_V (a)$; by compactness, $J_i(a) \cdot a$ is within a finite radius X/E-ball B around a; now B is finite by local finiteness; so a finite index definable subgroup of $J_i(a)$ fixes it pointwise. But $J_i(a)$ is connected, so $J_i(a)$ acts trivially on B; in particular $J_i(a)$ fixes the E-class of a. Let X' be the directed graph obtained by drawing a directed edge from a to $j \cdot a$, for any $j \in J_i(a)$. We have just shown that the forward neighbors of a are in the same E-class as a. Applying this to other elements, we see that the backwards-neighbors of a are also in the same E-class. But by assumption, any edge of X is in X' or the reverse graph; so all X-neighbors of a are in the same E-class. Conversely we have $E \subset \sim_V$, so the E-classes coincide with the $\sim_V$ classes. Hence both are definable. $\square$

For example we may take $\mathcal{G}$ to be the ind-definable group of polynomial automorphisms of $V = \mathbb{A}^n$. By a *transvection* (not necessarily linear) we mean a map $(x, y) \mapsto (x, y + f(x))$, where (x, y) is a partition of the coordinates of $\mathbb{A}^n$, and f is a polynomial. Note that any transvection t forms part of a unipotent group $(t^\alpha : \alpha \in F)$, where $t^\alpha(x, y) = (x, y + \alpha f(x))$; in the theory T of pseudofinite fields of characteristic zero, this group is connected.

Corollary 6.10. *Let j_1, j_2 be definable, conjugation invariant maps from V to $\mathcal{G}$, so that $j_i(a)$ is a transvection; draw an edge from a to $j_i(a) \cdot a$. Let Γ denote the associated graph (of valency at most 4). Then the connected component $\sim_V (a)$ in the finite field $\mathbb{F}_p$ is definable uniformly in a and in p.* $\square$

This is a special case of Corollary 6.9. The proof shows also that the graph (V, X') obtained by connecting a to each $j_i^t(a)$, $i \in I$, $t \in F$, has the same connected components as (V, X), and each connected component has finite X'-diameter.

Example 6.11. Fix a linear algebraic group G (say over $\mathbb{Z}$). Consider the action of G by conjugation on G^n. Define a graph structure Γ on G^n, of valency at most n, by letting a be adjacent to $a_i^{-1} a a_i$, where a_i is one of the n components of $a = (a_1, \dots, a_n)$ and $k \in \mathbb{Z}$. Let p be a prime, let $a_1, \dots, a_n \in G(\mathbb{F}_p)$ be unipotent, $a = (a_1, \dots, a_n)$, and let $C(p; a)$ be the component of a in $\Gamma(\mathbb{F}_p)$. Then $C(p; a)$ is definable in $\mathbb{F}_p$ uniformly in a and p.

If one "speeds up" the graph by declaring $a_i^{-k} a a_i^k$ to be adjacent to a, then there exists a bound β such that for all primes p and all a as above, $C(p; a)$ has diameter $\leq \beta$.

Each component of Γ is contained in a conjugacy classes of n-tuples of unipotent elements, in the subgroup of G that they generate. The subgroup $G(a)$ generated by an n-tuple a is itself a definable function; this is a consequence of Theorem 6.1. From Theorem 6.6 we learn that the component of a is a definable subset of the $G(a)$-conjugacy class of a. In particular when $G = \mathrm{SL}_r$ and a is a generating n-tuple of $\mathrm{SL}_r(\mathbb{F}_p)$, I do not know if the conjugacy class of a is connected in Γ; if it is not, the theorem shows that it is due to an algebraic invariant. In particular if the conjugacy class in $G(\mathbb{F}_p)^n$ is connected away from a density zero set of primes, then it is connected for all but finitely many primes.

See [15; 14; 1; 19] and references there for deep work on more specific instances or classes of definable graphs.

6.12. ***Example of unbounded first expansion radius.*** We conclude with a simple example showing that the "layered" or "tree-like" structure in the statement of Theorem 6.6 is unavoidable.

Let (G, X) be a (symmetric, reflexive) graph, definable in a structure of finite S1-rank $d(G)$. For $a \in G$, let $\xi(a)$ be the smallest n such that for all $n' > n$ we have $d(B_{n'}(a)) = d(B_n(a))$, where $B_n(a)$ is the n-ball of (G, d_X).

Here is an example where ξ is unbounded. It can be understood in ACF or in PF, and shows that the finite partition in Theorem 6.6 is unavoidable.

Example 6.13. Let $G = \mathbb{A}^2$, and let $f : G \to G$ be an endomorphism of infinite order. Let C be an irreducible curve on G, also of infinite order under f, i.e., not f-preperiodic. Let X be the union of the graph of f and $C \times C$. Then ξ is unbounded. Hence, there is no definable equivalence relation H on G such that:

(1) For some m, if $(a, b) \in H$ then $d_X(a, b) \leq m$.

(2) X/H has finite valency.

Indeed, $\xi(g) = n$ if $g \in f^{-n}(C) \setminus \bigcup_{k<n} f^{-k}(C)$. Thus ξ is unbounded. Suppose an H with (1,2) exists. Let $a \in G$ be generic; it suffices that $a \notin \bigcup_{n\in\mathbb{Z}} f^n C$. Since all d_X-balls around a are finite, and by (1), the H-class of a is finite. Hence all H-classes must be finite except possibly on some finite union of curves $f^{-n}(C)$. On the other hand, the curve C itself meets only finitely many H-classes, since the complete graph on C has finite valency modulo H. Let n be the greatest integer such that $f^{-n}(C)/H$ is finite. Then $f^{-n-1}(C)/H$ is infinite, yet f induces a finite valency graph on the product $f^{-n-1}(C)/H \times f^{-n}(C)/H$, which is impossible.

6.14. ***Larsen–Pink.*** One may ask about a graph-theoretic version of the Larsen–Pink inequalities. We obtain such an analogue in a basic case, leaving the more general case as a question.

Consider an enrichment of the theory of fields, with a reasonable dimension function δ, as in [46]. For any partial type S, write $d(S)$ for the dimension of the Zariski closure of S, also $d(c) = d(S)$ if $S = \mathrm{tp}(c/M)$. Write $\delta(S) = \alpha$ if $\alpha = \inf \delta(S')$ as S' varies over definable sets containing S.

Let P, Q and $R \subset P \times Q$ be complete types over some base field M, such that $\delta(P), \delta(Q), \delta(R)$ exist. Assume

$$\text{for } (a,b) \in R \text{ we have } M(a)^{\mathrm{alg}} \cap M(b)^{\mathrm{alg}} = M. \qquad (\diamond)$$

Lemma 6.15. *Assume $d(P) = d(Q)$ and $d(R) = d(P) + 1$. Then*

$$d(P)\delta(R) \le \delta(P)d(R).$$

Proof. Let $R^t = \{(b,a) : (a,b) \in R\}$. Consider $(a_0, \dots, a_k)$ with $(a_i, a_{i+1}) \in R \cup R^t$ such that $a_0 \mathop{\smile\!\!\!\!\!\!\downarrow}_{M(a_i)} a_{i+1}$. Then $d(a_0 a_{i+1}) \le d(a_0 a_i a_{i+1}) = d(a_0 a_i) + 1$. Thus $d(a_0 a_i) \le d(P) + i$. Once $d(a_0 a_i) = d(a_0 a_{i+1})$ we have $a_0 \mathop{\smile\!\!\!\!\!\!\downarrow}_{M(a_{i+1})} a_i$ and so by considering the canonical base we have $a_0 \mathop{\smile\!\!\!\!\!\!\downarrow}_{M(a_i)^{\mathrm{alg}} \cap M(a_{i+1})^{\mathrm{alg}}} M(a_i)^{\mathrm{alg}}$, i.e., by assumption, $a_0 \mathop{\smile\!\!\!\!\!\!\downarrow}_M a_i$; so $d(a_0 a_i) = 2d(P)$. Thus equality holds only at $i = d(P)$, and for $i < d(P)$ we have $a_i \in M(a_0 a_{i+1})^{\mathrm{alg}}$.

So far we used only algebraic independence. But now choose $(a_0, \dots, a_k)$ with $(a_i, a_{i+1}) \in R \cup R^t$ such that $a_0 \mathop{\smile\!\!\!\!\!\!\downarrow}_{M(a_i)} a_{i+1}$ holds in the sense of δ (and a fortiori algebraically). Using $a_i \in M(a_0 a_{i+1})^{\mathrm{alg}}$, we have $\delta(a_i / M(a_0 a_{i+1})) = 0$, and it follows that for $i \le d(P)$ we have $\delta(a_0 a_i) \ge \delta(a_0 a_{i-1}) + \delta(R_b)$, where $R_b = \{x : (x,b) \in R\}$. So $2\delta(a_0) \ge \delta(a_0 a_i) \ge \delta(a_0) + i\delta(R_b)$. For $i = d(P)$ we obtain $\delta(a_0) \ge d(P)\delta(R_b)$. Thus $\delta(R_b) \le \delta(P)/d(P)$ and also $\delta(R) = \delta(P) + \delta(R_b) \le \delta(P)(1 + 1/d(P))$ so $\delta(R)d(P) \le \delta(P)d(R)$. $\square$

Question 6.16. What can be said without the condition $d(R) = d(P) + 1$?

Let us now see what $(\diamond)$ amounts to for graphs arising from a definable subset X of a definable group G. First we analyze the condition $M(a)^{\mathrm{alg}} \cap M(b)^{\mathrm{alg}} = M$ in this case, showing that it means in essence that X generates G. The argument of this paragraph is valid in theories of finite Morley rank, in particular the case that concerns us, ACF. Let G be a connected definable group, $X \subset G$ a complete type over M. Let (b, c) be generic in $G \times X$, and let $a = bc$.

Claim. *We have* $\mathrm{acl}(M(a)) \cap \mathrm{acl}(M(b)) = M$ *iff X is not contained in an M-definable coset of a proper subgroup of G.*

Indeed if $X \subset Hm$ then $aHm = bHm \in \mathrm{acl}(M(a)) \cap \mathrm{acl}(M(b))$ but by genericity of b in G, $bHm \notin \mathrm{acl}(M)$. Conversely, assume $\mathrm{acl}(M(a)) \cap \mathrm{acl}(M(b)) \neq M$. Let $\mathrm{tp}(b'/\mathrm{acl}(M(a)) = \mathrm{tp}(b/\mathrm{acl}(M(a))$, with b', b independent over $M(a)$. Since $b = ac^{-1}$, there exists d^{-1} with (b, c, d) generic in $G \times X \times X$, and $b' = ad^{-1} = bcd^{-1}$. If $e \in \mathrm{acl}(M(a)) \cap \mathrm{acl}(M(b)) \setminus M$, we still have $e \in \mathrm{acl}(M(b'))$. Continuing this way, for arbitrarily large n we can find $(b, c_1, \dots, c_{2n})$ generic in $G \times X^{2n}$, with $\mathrm{acl}(M(bc_1c_2^{-1} \cdots c_{2n-1}c_{2n}^{-1})) \cap \mathrm{acl}(b) \neq M$. For large enough n, $c_1c_2^{-1} \cdots c_{2n-1}c_{2n}^{-1}$ is a generic of element of a definable group $H \leq G$; now $H = G$ is impossible since for generic $c \in G$, we do have $\mathrm{acl}(Mbc) \cap \mathrm{acl}(Mb) = M$. Thus $c_1c_2^{-1} \cdots c_{2n-1}c_{2n}$ is the generic of a proper definable subgroup, and the result follows.

Remark 6.17. Let G be a definable group and X a definable subset. Let $P = Q = G$ and let $R = \{(x, y) \in G^2 : xy^{-1} \in X\}$. Then the inequality

$$d(P)\delta(R) \leq \delta(P)d(R)$$

implies the Larsen–Pink inequality

$$\delta(X)d(G) \leq \delta(G)d(X).$$

To see this, renormalize δ so that $\delta(G) = d(G)$. Then the first inequality becomes $\delta(R) \leq d(R)$; equivalently $\delta(G) + \delta(X) \leq d(G) + d(X)$. Using $\delta(G) = d(G)$ additively now, we obtain $\delta(X) \leq d(X)$ and hence $\delta(X)d(G) \leq \delta(G)d(X)$.

7. The Galois group of a NIP measure

Let $\mu(x)$ be a definable measure in a NIP theory, let $\phi(x, u)$ be a formula, and $q(y)$ a type (over 0).

Define an equivalence relation on q as follows: $bE_{\mu,\phi}b'$ iff for μ-almost all x, $\phi(x, b) = \phi(x, b')$. This is then a $\bigwedge$-definable equivalence relation on q. In a NIP theory, ([45]), it is cobounded and thus the quotient q/E is compact in the logic topology. The group G of automorphisms of q/E is likewise a compact group. Since q is a complete type, G acts transitively on q/E. In this section we will study this group $G = G_{\mu,\phi,q}$.

We could also consider finitely many formulas $\phi_1(x, u_1), \dots, \phi_k(x, u_n)$, each with a type $q_i(u_i)$; but by standard tricks one can find a single type $q = q(u_1, \dots, u_n)$

projecting to q_i at the i-th position, and a formula ϕ, such that each $\phi_i(x, c_i)$ ($c_i \models q_i$) is equal to some $\phi(x, c)$ (with $c \models q$). Thus the sets of such formulas $\phi(x, c)$ with $c \models q$ forms a directed system, and the projective limit of the groups $G = G_{\mu,\phi,q}$ can be viewed as the Galois group of the space of weakly random types. The individual groups $G_{\mu,\phi,q}$ are the fundamental building blocks.

It turns out that a totally disconnected part may appear both as a quotient (G/G^0) and as a normal subgroup K/G^{00} of G^0/G^{00}. We show in Theorem 7.12 that the "archimedean core" G^0/K is a finite-dimensional real Lie group.

As a corollary we obtain Theorem 7.15, addressing a basic question of Anand Pillay regarding local finite-dimensionality of groups of connected components G/G^{00} for a definable group G in a NIP theory. We define the local connected component $G^{00}_{\phi,q}$ attributable to $\phi(x, u)$ ranging over a given type $q(u)$, and show again that it is a finite-dimensional Lie group, up to profinite group extensions above and below.

We first present some basic material characterizing the target, profinite-by-Lie-by-profinite groups. In this section, by "Lie group" we will mean finite-dimensional Lie group, with finitely many connected components. Only compact Lie groups will be considered. All topological groups are taken to be Hausdorff.

7.1. ***Profinite extensions of compact Lie groups.*** Let L be a connected, finite-dimensional Lie group, with identity element 1, and let $\pi_1(L) := \pi_1(L, 1)$. Then the universal cover $\tilde{L}$ admits a group structure, induced for instance by multiplication of paths; there is an exact sequence $1 \to \pi_1(L) \to \widetilde{L} \to L \to 1$. The connected group $\tilde{L}$ acts trivially by conjugation on the finitely generated group $\pi_1(L)$, so this is a central extension. Any subgroup N of $\pi_1(L)$ of finite index is normal in $\widetilde{L}$, and we can form the quotient $\widetilde{L}/N$. Taking the inverse limit over all N we obtain a group $\hat{L}$; by construction it is an extension of the Lie group L by the finitely generated profinite group $\widehat{\pi_1(L)}$, profinite completion of $\pi_1(L)$. $\hat{L}$ admits a surjective continuous map onto any connected, pointed, finite covering group of L (unique with $1 \mapsto 1$), and at the limit, a surjective continuous map onto any connected central extension E of L by a profinite group. We call $\hat{L}$ the *universal profinite covering group* of L.

7.2. ***Groups of finite archimedean rank.***

Definition 7.3. A connected compact topological group K is *profinite-by-Lie*, if it has a closed normal profinite subgroup K_0 such that K/K_0 is a Lie group.

We will say that compact topological group G is profinite-by-Lie-by-profinite, or *has finite archimedean rank*, if the connected component G^0 of G is profinite-by-Lie.

We will see that G is actually profinite-by-Lie-by-finite in this case, though G/G^0 need not be finite.

Remarks 7.4. (1) If K is a compact group, then K/K^0 is profinite.

(2) Let G be a topological group with a profinite normal subgroup K, such that G/K is a Lie group with finitely many connected components. Then K is determined up to finite index by these conditions. Indeed if L is another such group, then KL/K is a normal profinite subgroup of G/K; as Lie groups do not contain infinite normal profinite groups, KL/K must be finite, and by symmetry K, L are commensurable.

(3) If K_0 is profinite and normal in a connected group G^0 then it is central in G^0. Indeed the connected group G^0 acts by conjugation on K_0, so each orbit is connected, but the totally disconnected group K_0 has no connected subsets other than points.

(4) Let A be a compact abelian group; we have $A \cong \mathrm{Hom}(\hat{A}, U_1)$, where U_1 is the unit circle in $\mathbb{C}$, and $\hat{A}$ is the character group. Then A has finite archimedean rank iff $\hat{A}$ does as a discrete group, i.e., $\mathbb{Q} \otimes \hat{A}$ is a finite-dimensional $\mathbb{Q}$-vector space.

(5) A compact group G has finite archimedean rank iff the center Z of G has finite archimedean rank, and G/Z has a finite-index subgroup isomorphic to the product of a profinite group with a compact real Lie group. (See Proposition 7.6(3).)

Example 7.5. Take a finite-dimensional Lie group L, a central extension $\bar{L}$ of L by a profinite group A, another profinite group P, and a central extension $\bar{P}$ of P by A; and form $G = \bar{P} \times^A \bar{L}$, the quotient of $\bar{P} \times \bar{L}$ by the antidiagonal subgroup of A^2. Then G is profinite-by-Lie-by-profinite, and compact if L is.

In fact G is also profinite-by-Lie in this case, as shown by the image of $\bar{P} \times A$ in G. Along with Proposition 7.6(3), this explains the remark in brackets in Definition 7.3.

Proposition 7.6. *Let G be a compact topological group.*

(1) *If G is connected profinite-by-Lie, then G is a quotient of the universal profinite cover of a compact connected Lie group.*

(2) *If G is Lie-by-profinite, then some subgroup G' of finite index in G splits as a direct product of a connected Lie group with a profinite group, both normal in G.*

(3) *If G is of finite archimedean rank, then G has an open finite index subgroup isomorphic to $\bar{P} \times^A \bar{L}$, with $\bar{P}$, A, $\bar{L}$ as in Example 7.5. Hence G is profinite-by-Lie-by-finite.*

(4) *G has finite archimedean rank iff there exists a bound r and a family $(N_i : i \in I)$ of closed normal subgroups of G, closed under finite intersections, such that $\bigcap_{i \in I} N_i = (1)$ and each G/N_i is a Lie group of dimension $\leq r$.*

(5) *Let H be a closed subgroup of G. If G has finite archimedean rank, then so does H.*

Proof. (1) This follows from the discussion in Section 7.1 above.

(2) We have an exact sequence

$$1 \to L \to G \to P \to 1$$

with L a Lie group, P profinite. As L has no small subgroups, G has an open neighborhood O containing no nontrivial subgroup of L. By Peter–Weyl there exists a continuous homomorphism $r : G \to U_n \leq \mathrm{GL}_n(\mathbb{C})$ with kernel R contained in O and hence meeting L trivially. (See, e.g., [72, 1.4.14] for this consequence of Peter–Weyl.) On the other hand $G/RL \cong P/\operatorname{Im}(RL)$ is a profinite group, that embeds in a subquotient of $\mathrm{GL}_n(\mathbb{C})$; so it finite. Thus RL has finite index in G, and splits as a semidirect product of R with L. Conjugation induces a homomorphism

$$R \to \operatorname{Aut}(L) \to \operatorname{Aut}(\operatorname{Lie}(L)) \subset \mathrm{GL}_n(\mathbb{C}),$$

where $\operatorname{Lie}(L)$ is the Lie algebra of L. The composition $R \to \mathrm{GL}_n(\mathbb{C})$ has finite image as R is profinite; so some open subgroup R' of R acts trivially on $\operatorname{Lie}(L)$, hence on the connected component L^0; so a finite index normal subgroup R'' of R acts trivially on L. We may take R'' normal in G. Thus $R''L$ is a direct product as stated. Note that $R \cong RL/L$ is isomorphic to a subgroup of P, and hence is profinite.

(3) We have two exact sequences:

$$1 \to G^0 \to G \to P \to 1,$$
$$1 \to Q \to G^0 \to L \to 1,$$

with P, Q profinite, L Lie, G^0 connected. Q is contained in the center of G^0 (since a connected compact group has abelian π_1, and acts trivially by conjugation on it).

Since L has no small subgroups, there exists a neighborhood U of 1 in G such that any subgroup of G contained in $U \cap G^0$ is contained in Q. By Peter–Weyl there exists a homomorphism $r : G \to \mathrm{GL}_n(\mathbb{C})$ with kernel $R \subset U$. So $R \cap G_0 \leq Q$. $G/(RG^0)$ is a quotient of the profinite group G/G^0 and also of the group G/R that has finitely many connected components; so $G/(RG^0)$ is finite, i.e., RG^0 has finite index in G. We may thus assume $RG^0 = G$, and likewise replace Q by $R \cap G^0$, and L by the new G^0/Q; now Q is normalized by R.

Since the closed subgroup $R \cap G^0$ is commensurable with Q, it is profinite. Since R/G^0 is also profinite, R is profinite.

Since G^0 is connected, it acts trivially on the totally disconnected group R; i.e., G^0, R commute. Thus R acts trivially on G^0. (Thanks to the referee for this short argument.) In particular Q is central in $RG^0 = G$. So $G = R \times_Q G^0$.

(4) Assume such a bound exists. Then $\dim(G/N_i)$ is bounded independently of i. It follows that for some N_0, for all $N_i \leq N_0$, $\dim(G/N_0) = \dim(G/N_i)$; so

$[N_0 : N_i] < \infty$. Since $\bigcap\{N_i : N_i \le N_0\} = 1$, N_0 is profinite; while G/N_0 is a compact Lie group.

Conversely, assume G has finite archimedean rank. By (3), G is profinite-by-Lie-by-finite, so it has an open profinite-by-Lie normal subgroup G_1 of finite index; and clearly G_1 has a family $(N'_i : i \in I)$, with $\dim(G_1/N'_i) \le r'$ and $\bigcap_i N'_i = 1$. Let N_i be the intersection of the conjugates of N'_i. Each N'_i has at most $[G : G_1]$ conjugates, and thus G/N_i is a Lie group of dimension $\le r'[G : G_1]$; clearly $\bigcap_i N_i = (1)$.

(Thus G is profinite-by-(not necessarily connected Lie).)

(5) This follows from (4), since a closed subgroup of a Lie group is a Lie group of lower (or equal) dimension. □

We actually need the following lemma only when L is compact, but state it more generally with a view to a later generalization (see Section 7.18(2)).

Lemma 7.7. *Let L be a Lie group acting transitively and faithfully on a connected manifold Y, with compact stabilizer S of a point $y \in Y$. Then* $\dim(L) \le \dim(Y)^2$.

Proof. Since S is compact, fixing $y \in Y$ one can find an S-invariant inner product b on the tangent space T_y, and this propagates to an L-invariant Riemannian metric on Y. Using the exponential map one sees that the pointwise stabilizer S' of the tangent space T_y fixes a neighborhood of y. The set of points y' such that y' and each vector in $T_{y'}$ is fixed by S' is closed and open, hence equals Y. So $S' = 1$, hence the homomorphism $L \to \operatorname{End} T_y$ is injective. Thus $\dim(S) \le \operatorname{Aut}(T_y, b) = \binom{\dim Y}{2}$, and hence

$$\dim(L) \le \dim(Y) + \binom{\dim Y}{2} = \frac{1}{2}(\dim(Y)^2 + \dim(Y)) \le \dim(Y)^2. \qquad \square$$

Let Y be a compact C^1 differentiable manifold. Let d_Y be a metric on Y. We say that d_Y is *compatible with the manifold structure* if for any chart $c : W \to Y$, where W is an open subset of $\mathbb{R}^n$ and d_W is the metric induced from $\mathbb{R}^n$, the map $c : (W, d_W) \to (Y, d_Y)$ is bi-Lipschitz. Any two compatible metrics on Y are bi-Lipschitz equivalent, and hence the packing dimension is well-defined and does not depend on the choice of a compatible metric. In fact the packing dimension equals the dimension of Y as a manifold, as can be seen by going to charts and using the standard Euclidean metric.

Lemma 7.8. *Let G be a compact Lie group, H a finite-dimensional Euclidean space, $\rho : G \to \operatorname{Aut}(H)$ a faithful unitary representation. Let $X = Gv$ be an orbit of G, with metric d_X induced from H. Then X is a smooth submanifold of H, and d_X is Lipschitz-compatible with the manifold structure on X.*

Proof. Let F be the graph of $\rho : G \to \operatorname{Aut}(H)$. Then F is a closed subgroup of $G \times \operatorname{Aut}(H)$ which is analytic by the closed subgroup theorem (von Neumann 1929, Cartan 1930). Since the projection $F \to G$ is analytic, and an isomorphism, the inverse map $G \to F$ is too; composing with the projection $F \to \mathrm{Aut}(H)$ we see that ρ is a real-analytic, and in particular C^1, isomorphism between G and an analytic subgroup A of $\mathrm{Aut}(H)$.

$G_v = \{g \in G : gv = v\}$. We give G/G_v the quotient manifold structure [13, 5.9.5]. Likewise let $A_v = \{a \in A : av = v\}$, and give A/A_v the quotient manifold structure. We obtain induced analytic isomorphisms $G/G_v \to A/A_v$.

Consider the map $\alpha : A \to X$, $a \mapsto a(v)$, and the derivative α' of α at 1, from the Lie algebra of A into $T_v H = H$. The kernel is precisely the Lie algebra $T_1 A_v$ of A_v. Hence the induced map $A/A_v \to H$ induces an injective linear map on tangent spaces, and so is an immersion by [13, 5.7.1]. Since $A/A_v \cong G/G_v$ is compact and $A/A_v \to H$ is injective, it follows that the image X is a differentiable submanifold of H, and $G/G_v \cong A/A_v \to X$ is an isomorphism of C^1-manifolds.

For the final point, it is easy to see more generally that if Y is a smooth submanifold of $\mathbb{R}^n$, the metric induced by a Euclidean structure on $\mathbb{R}^n$ is compatible with the manifold structure on Y. For instance near a point $a \in Y$, let H' be the tangent space to Y at a embedded as a linear subspace of H, and consider the orthogonal projection $\pi : Y \to H'$; at an ϵ-neighborhood of a the distance to H is $O(\epsilon^2)$, and so π is bi-Lipschitz. □

Lemma 7.9. *If $f : Y \to X$ is a Lipschitz map between metric spaces, then the packing dimension of Y is at least equal to the packing dimension of X.*

Proof. In fact let $\gamma_X(\epsilon)$ be the maximal number of disjoint ϵ-balls in X, likewise $\gamma_Y(\epsilon)$, and let l be the Lipschitz constant of f; then $\gamma_Y(\epsilon) \geq \gamma_X(l\epsilon)$. If $a_1, \dots, a_k$ are points of X such that the $l\epsilon$-balls around the a_i are pairwise disjoint, let $b_i \in X$ be such that $f(b_i) = a_i$. Then the ϵ-balls around the b_i are pairwise disjoint: if $c \in Y$ and $d_Y(b_i - c), d_Y(b_j - c) < \epsilon$, and $d = f(c)$, then $d_X(a_i, d), d_X(a_j, d) < l\epsilon$, a contradiction to the disjointness of the $l\epsilon$-balls around a_i, a_j. □

Lemma 7.10. *Let G be a compact, connected Lie group, H a Hilbert space, $\rho : G \to \mathrm{Aut}(H)$ a unitary representation. Let X be an orbit of G, and assume G acts faithfully on X. Let δ be the packing dimension of X as a metric space with the metric induced from H. Then $\dim(G) \leq \delta^2$.*

Proof. By Peter–Weyl, there exist finite-dimensional G-invariant subspaces H_i' ($i \in \mathbb{N}$) whose direct sum is dense in H. Let $H_n = \bigoplus_{i \leq n} H_i'$, and $\pi_n : H \to H_n$ the orthogonal projection. Since π_n is Lipschitz (in fact with constant 1), $\pi_n(X)$ still has packing dimension $\leq \delta$. Also, if $g \in G$ fixes $\pi_n(x)$ for each n and each point $x \in X$, then g fixes X and hence H, so $g = 1$. Let $K_n = \{g \in G : g|\pi_n(X) = \mathrm{Id}\}$. Then $\bigcap K_n = (1)$. By the descending chain condition on closed subgroups of G we see that $K_n = 1$ for some n, so the action of G on $\pi_n(X)$ is faithful. Replacing H by the subspace of H_n generated by $\pi_n(X)$, and using Lemma 7.9, we are reduced to the case that H is finite-dimensional.

In this case, by Lemma 7.8, δ equals the dimension $\dim(X)$ of X as a manifold. By Lemma 7.7, $\dim(G) \le \dim(X)^2 \le \delta^2$. $\square$

7.11. ***Interaction of NIP formulas with definable measures.*** We consider a sort X carrying a definable measure, and a NIP formula $\phi(x, u)$ relating another sort U to X. More generally we allow U to be a $\bigwedge$-definable set (without parameters); and the measure μ may just be defined on the algebra $B(\phi)$ of subsets of X generated by the sets $\phi(x, b)$, $b \in U$.

We will find a compact Hausdorff quotient U of U through which the interaction is mediated. We are interested in the fibers of the natural map from U to the space of types over $\varnothing$. These are principal homogeneous spaces for a compact group G. We show here that G is made up of totally disconnected groups and a single finite-dimensional Lie group.

We begin by recalling a number of basic objects associated with (X, U, ϕ, μ): a compact space X with a probability measure, a compact metric space U, a compact topological group G acting on X and U.

Let $M = (X, U, \phi, \mu)^M$ be an $\aleph_1$-saturated model. We assume $\mathrm{Aut}(M)$ acts transitively on $U(M)$ (i.e., we consider one type of U at a time).

We can define a pseudometric on U:

$$d(a, a') = \mu(\phi(a, x) \triangle \phi(a', x)),$$

with associated equivalence relation

$$a E_{\phi,\mu} a' \iff \mu(\phi(a, x) \triangle \phi(a', x)) = 0.$$

So the quotient $\mathsf{U} = \mathsf{U}_{\phi,\mu} = U/E_{\phi,\mu}$ becomes a metric space.

$E = E_{\phi,\mu}$ is a $\bigwedge$-definable equivalence relation. Clearly $a E_{\phi,\mu} a'$ iff for all weakly random p over M, $\phi(a, x) \in p \iff \phi(a', x) \in p$. Since ϕ is NIP, the number of weakly random ϕ types has cardinality bounded independently of M (see Section 2.20). Hence E is cobounded. Thus we also have a logic topology on U. The identity map is continuous from the (compact) logic topology to the (Hausdorff) topology induced by the metric, so they coincide.

By compactness, for any $\epsilon > 0$ there is a finite bound on the size of a set $A \subset U(M)$ such that $\mu(\phi(a, x) \triangle \phi(a', x)) \ge \epsilon$ for all $a \neq a' \in A$. A maximal finite set of this kind can be found in $U(M_0)$ for any countable elementary submodel M_0 of M. It follows that U is separable, and does not depend on M. Being a compact metric space, U is second countable.

Let X be the space of weakly random ϕ-types.

We have a map $i : U \to L^2(X)$, mapping a to the characteristic function of $\phi(a, x)$. We have $\|i(a) - i(b)\|_2^2 = d(a, b)$. The definable measure μ induces a regular Borel measure on X, and i extends to an embedding $i : \mathsf{U} \to L^2(\mathsf{X})$.

Let G be the group of permutations of U induced by automorphisms of M. These are the permutations preserving all images in U of $\bigwedge$-definable subsets of U^n. In particular d is preserved, so G consists of isometries of U. We give G the topology of pointwise convergence (equivalently here, the uniform topology). It follows that G too is second countable. Since G is closed in the group of all isometries of U, it is compact. We will refer to $\mathsf{G} =: \mathsf{G}_{\mu,\phi}$ as the compact symmetry group of μ relative to ϕ, U.

The action of G extends canonically to an action by automorphism on X. As G preserves the relations giving the measure of any finite Boolean combination of $\phi(a, x)$, this action is measure-preserving. It thus induces a unitary action of G on $L^2(\mathsf{X})$. Note also that $g \mapsto (gu, v)$ is continuous. If an automorphism of M fixes U, it fixes X too, since it fixes the image of the relation ϕ on $\mathsf{U} \times \mathsf{X}$.

Conversely if g fixes X, it is clear from the definition of E that g fixes U.

Theorem 7.12. *Let $\phi \subset X \times U$ be a NIP formula, and assume U is a complete Shelah strong type, i.e., a $\bigwedge$-definable set carrying no nontrivial definable equivalence relations with finitely many classes. Let μ a definable Keisler measure on X.*

Then $\mathsf{G}_{\mu,\phi}$ has finite archimedean rank.

Proof. Let $\mathsf{G} = \mathsf{G}_{\mu,\phi}$, and let G^0 be the connected component of G. By the completeness assumption on U, G^0 acts transitively on U.

By Peter–Weyl there is a cofinal system $\mathcal{N}$ of closed normal subgroups N of G^0 such that G^0/N is a connected Lie group. By "cofinal" we mean that any neighborhood of the identity element of G^0 contains some $N \in \mathcal{N}$; equivalently, we have $\bigcap \mathcal{N} = (1)$. Since G is second-countable, we may take $\mathcal{N}$ to be countable.

For $N \in \mathcal{N}$, factoring out $(\mathsf{U}, \mathsf{X}, i, \mu)$ by the action of N yields $(\mathsf{U}/N, \mathsf{X}/N, i, \mu_N)$, where μ_N is the pushforward measure on X/N. We will also use the orthogonal projection $\pi_N : L^2(\mathsf{X}) \to L^2(\mathsf{X})^N$, where $L^2(\mathsf{X})^N$ is the subspace of $L^2(\mathsf{X})$ consisting of N-invariant functions. We have a canonical identification of $L^2(\mathsf{X})^N$ with $L^2(\mathsf{X}/N)$. By computing $\int fg$ for $g \in L^2(\mathsf{X}/N)$, it is easy to see that $\pi_N(f)$ is the integral of f with respect to Haar measure on N, i.e., $\pi_N(f)(x) = \int f(n(x))\, d_N(n)$. Hence, for continuous f on X, we have $f = \lim_N \pi_N(f)$, i.e., for any $\epsilon > 0$, for some open neighborhood W of 1 in G, whenever $N \subset W$, we have $\|\pi_N(f) - f\|_\infty < \epsilon$.

The action of G/N on $L^2(\mathsf{X}/N)$ is faithful: if $g \in G \setminus N$, then by Urysohn's lemma there exists a continuous function f on G/N vanishing at $1_{G/N}$ but not on g; clearly ${}^g f \neq f$.

Define $i_N = \pi_N \circ i : \mathsf{U} \to L^2(\mathsf{X}/N)$.

By Proposition 3.27, the image $i(\mathsf{U})$ of U in $L^2(\mathsf{X})$ has finite packing dimension. By Lemma 7.9, the packing dimension of $i_N(\mathsf{U})$ is at most δ, for each $N \in \mathcal{N}$. Since G^0 acts transitively on U, the image of U in $L^2(X/N)$ forms a single orbit of G^0/N-orbit. By Lemma 7.10, $\dim(\mathsf{G}^0/N) \leq \delta^2$. So $\dim(\mathsf{G}^0/N)$ is bounded, independently of $N \in \mathcal{N}$. By Proposition 7.6(4), G^0 has finite archimedean rank, hence (by definition) so does G. □

Remark 7.13. The theorem is valid more generally for continuous logic; the measure μ is better interpreted as an expectation operator in this case, but by Riesz, induces a measure μ on X; we still have an embedding of U into $L^1(X, \mu)$ mapping a to the real-valued function $\phi(a, x)$; we use the induced norm from L^1.

7.14. ***The local connected component of a NIP group.*** In this paragraph we address Pillay's question, discussed in the introduction.

Let G be a definable group. Let $\phi(x, u)$ be a NIP formula, with x ranging over G and u over a definable set U.

Assume G carries a definable, left translation invariant measure, at least on the Boolean algebra of subsets generated by left translates of formulas $\phi(x, b)$.

Recall the definition of a weakly random type, Section 2.20. Let G^{00}_{ϕ} be the intersection of all the stabilizers of weakly random ϕ-types. Equivalently,

$$g \in G^{00}_{\phi} \iff \mu(\phi(gx, b) \triangle \phi(x, b)) = 0 \text{ for all } b.$$

More generally, if $q(u)$ is a partial type, let

$$G^{00}_{\phi,q} = \{g \in G : \mu(\phi(gx, b) \triangle \phi(x, b)) = 0 \text{ for all } b \models q\}.$$

Thus G^{00}_{ϕ} is a $\bigwedge$-definable subgroup of G of bounded index, with compact quotient K. When $\phi(x, y) := (xy^{-1} \in P)$, P being a definable subset of G, we also write G^{00}_{P}.

We can view G/G^{00}_{ϕ} as the maximal quotient of G/G^{00} attributable to ϕ, and $G/G^{00}_{\phi,q}$ as the maximal compact quotient of G attributable to instances $\phi(x, b)$ with $b \models q$.

Theorem 7.15. *Let P be a definable subset of G. The compact group G/G^{00}_P has finite archimedean rank. More generally, if q is a complete Shelah strong type, then $G/G^{00}_{\phi,q}$ has finite archimedean rank.*

Proof. This reduces to Theorem 7.12 by a standard transposition from definable groups to semidirect factors of automorphism groups. Namely introduce a new sort X, with a regular action of G on X "on the right", and no additional structure beyond the original structure on the original sorts. We can view X as another copy of G; the new language includes the old and the map $X^2 \to G$, $(x, y) \mapsto x^{-1}y$. Define ϕ on X^2: $\phi(x, y) \iff P(x^{-1}y)$. The left-invariant definable Keisler measure μ induces a definable measure μ_X on X.

Let $\hat{G}$ be the compact symmetry group of μ_X relative to ϕ, as defined in Section 7.11. Now G acts by automorphisms on X: $(g, x) \mapsto gx$, fixing the old sorts including G itself. Indeed this identifies G with the automorphism group of the new structure over the old. It is easy to check that under this identification, G/G^{00}_P becomes a closed subgroup of $\hat{G}$. Hence by Theorem 7.12 and Proposition 7.6(5), G/G^{00}_P has finite archimedean rank.

The more general statement is deduced in the same way from Theorem 7.12, letting $U = X \times q$, and $\phi'(x, x', u) = \phi(x^{-1}y, u)$. Note that G acting on the right is transitive on X while fixing q pointwise, while for homogenous $M \models T$, $\mathrm{Aut}(M)$ is transitive on q, so altogether the automorphism group of the new structure is transitive on U. $\square$

Example 7.16. Take G to be a saturated elementary extension of $(\mathbb{Z}_p, +, D)$, where $D = \{x \in \mathbb{Z}_p : v_p(x) \in 2\mathbb{Z}\}$. Then $G^0 = G^{00}$. The measure-theoretic stabilizer of D is G^0, and $G/G^0 \cong \mathbb{Z}_p$.

Example 7.17. Here we show that the Lie group can interact with the totally disconnected group in Theorem 7.15, so that G^0/G^{00} is not a product of a totally disconnected group with a Lie group. Note that $\mathbb{Z}(1, 1)$ is a discrete subgroup of the topological group $\mathbb{Z}_p \times \mathbb{R}$; it intersects $\mathbb{Z}_p \times (-1, 1)$ trivially. Let $G = (\mathbb{Z}_p \times \mathbb{R})/\mathbb{Z}(1, 1)$. The pathwise connected component of the identity in G is a dense subgroup, the image of $\mathbb{R}$, and so G is connected. The image of $\mathbb{Z}_p$ is a compact normal subgroup of G, with $G/\mathbb{Z}_p \cong T := \mathbb{R}/\mathbb{Z}$, so G is profinite-by-Lie.

It is easy to see that the subgroup generated by $(1, 1)$ is relatively p-divisible in $(\mathbb{Z}_p \times \mathbb{R})$, and G has no nonzero p-torsion elements. Thus the exact sequence $0 \to \mathbb{Z}_p \to G \to \mathbb{R}/\mathbb{Z} \to 0$ does not split; even as a pure group, G does not contain a copy of $T = \mathbb{R}/\mathbb{Z}$, and this will not be fixed by moving to a subgroup of finite index, or a finite quotient.

As a set, G can be identified with $\mathbb{Z}_p \times [0, 1)$. The group law is then defined by

$$(z, t) + (z', t') = (z + z' + c(t, t'), m(t, t')),$$

where $c(t, t') \in \{0, 1\}$, and $0 \le m(t, t') < 1$ in $\mathbb{R}$. Thus the group G is definable in the ring $\mathbb{Q}_p \times \mathbb{R}$ (equivalently, in the model-theoretic disjoint union of the valued field $\mathbb{Q}_p$ and the ordered field $\mathbb{R}$). The theory of this structure is NIP.

The product measure of the two Haar measures on $\mathbb{Z}_p$ and on $[0, 1)$ (identified with $\mathbb{R}/\mathbb{Z}$) is invariant for this multiplication too. So it gives a generically stable measure. Let $D = \left\{(x, y) : v(x) \in 2\mathbb{Z},\ y \in \left[0, \tfrac{1}{2}\right)\right\}$, say: a definable set of positive measure.

Let M^* be a saturated elementary extension. Let S be the μ-stabilizer of D. It's a $\bigwedge$-definable subgroup of bounded index, and $G(M^*)/S \cong G$. G is connected and profinite-by-Lie, but cannot be split as a product.

7.18. ***Open questions.***

(1) Is the transitivity assumption on U necessary in Theorem 7.12? The original question was for *all* $\phi(x, b)$.

The question reduces to U comprising finitely many types $U_1, \dots, U_k$, provided the bound on $\dim(G)$ does not depend on k. The issue is that the action on each U_i may have a kernel N_i. For definable N_i, the Baldwin–Saxl lemma would imply that the intersection of k_0 of the N_i is already trivial, where k_0 is bounded in terms of the VC-dimension. Since N_i is only $\bigwedge$-definable, Baldwin–Saxl does not directly apply and some substitute is needed. In any case we obtain that the compact group G embeds into a product of Lie groups of bounded dimension; in case μ takes only finitely many values, the μ-stabilizer of each formula $\phi(x, b)$ is definable, and so the usual Baldwin–Saxl applies and shows transitivity is not needed.

(2) A local logic version, allowing for noncompact definable groups over $\mathbb{R}$ and $\mathbb{Q}_p$, would be very interesting.

(3) In [62], Macpherson and Tent study profinite definable groups G in a NIP structure M, along with a formula $\phi(x, u)$ such that any open subgroup has the form $\phi(x, b)$ for some b.[7] In this situation the Haar measure yields a natural definable measure on ϕ-types, and the "fullness" assumption implies that $G^{00} = G^{00}_\phi$ and profiniteness of $G(M)$ implies that $G/G^{00} \cong G(M)$. They show that $G(M)$ is a finite product of finite-dimensional p-adic analytic groups.

This can be seen as a profinite/adelic analogue of Pillay's conjecture on the archimedean part of G^{00}_ϕ, though under a much stronger hypothesis.

It would be very interesting to put these results on a common footing. In particular, does Theorem 7.12 have an adelic analogue? If G is a pro-p-group, is it p-adic analytic? Is G in general interpretable in the adeles, or rather in the disjoint union of finitely many p-minimal p-adic fields $\mathbb{Q}_p$ and the o-minimal field $\mathbb{R}$ (all possibly enriched analytically)?

Appendix A. Stability for invariant relations

We develop the basic results of stability, presented here in Theorems A.14 and A.27. We view them as a reduction, modulo a certain ideal, of binary relations to unary ones; thus a kind of measurability result for binary relations for the product measure. The theory is primarily due to Shelah, and for the most part we follow standard presentations. Shelah understood the significance of having the theorem over an arbitrary base structure and not just over an elementary submodel, and introduced imaginary elements and the algebraic closure as the precise obstructions to this. In [71], the theory was extended beyond the first-order setting. In [54], the main theorem was proved for arbitrary invariant stable relations *over a model*. A little later,

[7] Actually their formulation is the slightly stronger condition that the $\phi(x, b)$ for b from M define exactly the set of open subgroups. However it is probably sufficient for the main theorem of [62] to assume that the family contains all the open subgroups, as well as other groups that may have infinite index.

for simple theories, the "bounded closure" with its compact automorphism group was recognized by these authors as the obstacle to existence of 3-amalgamation. This was the first use of Lascar's compact Lascar group; in the case of finite S1-rank, 3-amalgamation was known to hold with ordinary algebraic closure and the associated profinite group (in [40]).

See [20] for a good presentation of the compact and general Lascar types; we will use it below. In [8], the theory was beautifully developed for continuous real-valued relations; Theorem A.27 is a (less elegant) generalization for more general $\bigwedge$-definable stable relations.

Here we treat arbitrary automorphism-invariant stable relations, over any base set. We show that the fundamental theorems of stability theory hold, with strong Lascar types as the natural obstacles to both uniqueness and existence.

For $\bigwedge$-definable relations, generalizing slightly the continuous real-valued case, we show that compact Lascar types or Kim–Pillay types suffice.

We begin with describing a local setting, allowing notably to discuss stable independence over an "imaginary" element of the form a/E, where E is a $\bigvee$-definable equivalence relation. We will need it in order to treat approximate equivalence relations canonically, in particular preserving any group actions on them. This generalizes the usual setting if one takes the metric d to be bounded.

We will sometimes assume the language is countable; the generalization to the general case (by considering countable sublanguages) is routine. A will denote a countable base set; we will sometimes use a countable elementary submodel M containing A. *The unqualified words definable*, $\bigvee$*-definable*, $\bigwedge$*-definable always mean*: *without parameters.*

When $R \subset X \times Y$, and $a \in X$, we let $R(a) = \{b : (a, b) \in R\}$. Define $R^t \subset Y \times X$, $R^t = \{(b, a) : (a, b) \in R\}$. When the context leaves no room for doubt, for $b \in Y$ we will write $R(b)$ for $R^t(b)$.

A.1. ***Local structures.*** Let $\mathbb{U}$ be a structure with a metric $d : \mathbb{U}^2 \to \mathbb{N}$, such that for any n, $\{(x, y) : d(x, y) \leq n\}$ is $\bigvee$-definable. Recall that a *definable* subset of $\mathbb{U}^n$ is the interpretation of a formula (without parameters); a $\bigvee$-definable set is a union of definable sets.[8] A typical way to obtain such a structure is to begin with an arbitrary binary relation R_0 on another structure $\mathbb{U}_0$. Let $\widetilde{E}$ be the equivalence relation generated by R_0. Then any $\widetilde{E}$-class is naturally a local structure; the metric distance $d(x, y)$ is the length of a shortest chain $x = x_0, \dots, x_n = y$ with $R(x_i, x_{i+1})$ or $R(x_{i+1}, x_i)$ for each i. Here the small distance relations are 0-definable; if one takes a family of relations instead, it is only $\bigvee$-definable.

[8] If many sorts are allowed, we still assume the domain of d is the set of all pairs, belonging to the union of all sorts. There are natural generalizations to bigger semigroups than $\mathbb{N}$, both in the direction of continuous metrics and of uncountable languages, but we restrict here to the main case.

A *graph* is a set Ω with a symmetric binary relation R. Let $R(a) := \{b : R(a, b)\}$. We define the associated *metric*

$$d_R(x, y) = \min\{n : (\exists x = x_0, \dots, x_n = y) R(x_i, x_{i+1}) \text{ for each } i < n\}.$$

The graph is connected if $d_R(x, y)$ is always defined. Note $R(a) \cup \{a\}$ is the d_R 1-ball around a. We define $R^k(a)$ to be the d_R k-ball around a, i.e., R^k is the composition of R with itself k times.

A relation $R(x_1, \dots, x_n)$ is *local* if for each $i, j \leq n$, for some m, $R(x_1, \dots, x_n)$ implies $d(x_i, x_j) < m$. (For unary relations, this poses no constraint.) We will be concerned only with local relations. We will say, when only local relations are allowed, that the structure is local. (This is closely related to the Gaifman graph, used in finite model theory, and to Gaifman's theorem on this subject.)

Note that one cannot freely introduce dummy variables; if we wish to involve an additional variable y, it must be added along with a formula that ensures $d(x_i, y) < m$ for some m. Geometrically this means we allow bounded products of the form $X \times_\delta Y = \{(x, y) : x \in X, y \in Y, \psi(x, y)\}$ where ψ implies $d(x, y) < \delta$ for some δ. Formulas are formed by such controlled addition of dummy variables, conjunction, disjunction or difference of formulas with the same set of variables, projections distance-bounded universal quantifiers, of the form $(\forall x)(d(x, y) \leq l \to \phi(x, y))$.

If E is a $\bigvee$-definable equivalence relation in a saturated structure, then each E-class can be presented as a local structure; the local structures setting will enable us to speak about independence over an E-class (viewed as a (generalized) imaginary element of the base). We can present E as having the form $d(x, y) < \infty$, where d is a metric such that $d(x, y) \leq n$ is definable, for each n. Then we can take the basic relations to be the d-bounded ones (this does not depend on the choice of d).

Any relation R has local traces $R_{|l}$, the intersection of R with distance $\leq l$ between any pair of variables. Note that R can be recovered from the $R_{|l}$, in the specific structure at hand; so that the automorphism group of the local structure obtained in this way is identical to the original one.

If a local structure $\mathbb{U}$ has a constant symbol, or more generally a nonempty bounded definable set D, then it can be viewed as a $\bigvee$-definable structure; it is the union of the definable sets of points at distance $\leq n$ from D, each of these being 0-definable. In general however, the automorphism group here need not respect any specific inductive presentation.

The metric can be extended to imaginary sorts: first to $\mathbb{U}^n$ via

$$d((x_1, \dots, x_n), (y_1, \dots, y_m)) = \max(\max_i \min_j d(x_i, y_j), \max_j \min_i d(x_i, y_j)),$$

then to a quotient by a bounded equivalence relation with quotient map $\pi : \mathbb{U}^n \to \mathbb{U}^n/E$ with distance defined by $d(u, v) = \inf\{d(x, y) : \pi(x) = u, \pi(y) = v\}$.

We assume $\mathbb{U}$ is saturated as a local structure, i.e., any d-ball is saturated; equivalently any small family of definable sets has nonempty intersection, provided the family includes a bounded set, and that any finite subset has nonempty intersection. Local saturation can be achieved by taking an ultrapower using bounded functions only.

A remark on ultraproducts: if (N_i, d_i) are a family of local structures for the same language, and (N, d) is an ultraproduct in the usual sense, one has an equivalence relation defined by: $d(x, y) \leq n$ for some standard n. Each equivalence class is a local structure, and Łos's theorem holds. Thus an ultraproduct here requires a choice of an ultrafilter along with a component, rather than just an ultrafilter.

A.2. *Locally compact Lascar types.* An $\bigwedge$-definable relation $E = \bigcap_i E_i$ is a *cobounded equivalence relation* if in any elementary extension N of the given structure, $\bigcap_i E_i(N)$ is an equivalence relation $E(N)$, and $N/E(N)$ has cardinality bounded independently of N.

Call a sort S *separated* if it carries an $\bigwedge$-definable cobounded local equivalence relation. If S is separated, let $\equiv^S_{\text{lc}}$ be the intersection of all $\bigwedge$-definable cobounded local equivalence relations on S. Then $\equiv^S_{\text{lc}}$ is the unique smallest such relation. It may change if we add parameters to the language. If the identity of S is clear we write simply $\equiv_{\text{lc}}$.

Let $\pi = \pi^{\text{lc}}_S : S \to S/\equiv_{\text{lc}}$ be the quotient map. On $S/\equiv_{\text{lc}}$ we define a topology: Y is closed iff $\pi^{-1}Y$ is locally $\bigwedge$-definable.

Lemma A.3. *The quotient by $\equiv_{\text{lc}}$ is a locally compact space* (*and σ-compact*) *space.*

Proof. (See [39; 20] for the bounded case, of Kim–Pillay spaces.) Let $a \in S$, and let B_n be the ball of radius n around a, in S. Then $\pi(B_n)$ is compact (so $S/\equiv_{\text{lc}}$ is σ-compact). Since $\equiv_{\text{lc}}$ is local, say $d(x, y) < m$ for $(x, y) \in S^2$ with $x \equiv_{\text{lc}} y$. Then $\pi(a) \notin \pi(S \smallsetminus B_m)$. But $\pi(B_m) \cup \pi(S \smallsetminus B_m) = \pi(S)$. Thus the compact set $\pi(B_m)$ contains a neighborhood of $\pi(a)$, namely the complement of $\pi(S \smallsetminus B_m)$. The proof of the Hausdorff property is similar. □

Remark A.4. The local algebraic closure acl(∅) in a given sort S can be defined as the union of the locally finite definable sets. The automorphism group of $\mathbb{U}$ has a quotient group acting faithfully on acl(∅), referred to as locally profinite Galois group of S. It is a totally disconnected locally compact group. The stabilizer of a nonempty subset of acl(∅) is a compact group (fixing one point implies leaving invariant balls of various radii).

One can similarly define the *local compact closure* to be the union of $S/\equiv_{\text{lc}}$, over all sorts S such that $\equiv_{\text{lc}}$ is defined.

Now consider the more general setting of Aut($\mathbb{U}$)-invariant equivalence relations (we will simply say: invariant to mean Aut($\mathbb{U}$)-invariant). Assume S has an Aut($\mathbb{U}$)-invariant cobounded local equivalence relation. Then it has a smallest one; it is denoted $\equiv_{\mathrm{Las}}$. This equivalence relation is generated by $\bigcup_m \theta_m(a,b)$, where $\theta_m(a,b)$ holds iff a,b begin an indiscernible sequence, and $d(a,b) \leq m$. When d has the property that any two elements are connected by a chain of elements of distance 1, as is the case in the main examples, $\equiv_{\mathrm{Las}}$ is generated by θ_2. At any rate, $\equiv_{\mathrm{Las}}$ is an F_σ relation (a countable union of $\bigwedge$-definable relations).

A.5. Assume

$$\text{for some } m_0, \text{ for all } n, \text{ any } n\text{-ball is a finite union of } m_0\text{-balls.} \qquad (\clubsuit)$$

Remark. The assumption $(\clubsuit)$ is true in the setting of a measure, finite on balls. More precisely assume μ is a definable measure, each ball of radius 1 has nonzero measure, and each ball of radius ≤ 3 has finite measure. Then by Rusza's trick, any ball of radius 3 is a union of finitely many balls of radius 2 (consider a maximal disjoint set of radius 1-balls in the radius 3 ball; then enlarging them to radius 2 would cover the larger ball). Assume in addition that the metric space is "geodesic" in the sense that any two points of length n are joined by a path of length n, where the successive distance is 1 (as is the case for Gaifman graphs). Then it follows inductively that any ball of radius n is a union of finitely many balls of radius 2.

Remark. We are interested only in types of elements at finite distance from elements of $\mathbb{U}$. In the presence of $(\clubsuit)$, any such type has bounded distance $\leq m_0$ from some element of $\mathbb{U}$. It follows that if $\mathcal{X}$ is an Aut($\mathbb{U}$)-invariant closed set of types over $\mathbb{U}$, then $\mathcal{X}$ contains a compact subset X with $\mathrm{Aut}(\mathbb{U})X = \mathcal{X}$ (namely the types of distance $\leq m_0$ from a given point).

A.6. ***Ideals of definable sets.*** We will work with saturated (local) structures $\mathbb{U}$. *Invariance* refers to the action of Aut($\mathbb{U}$), or Aut($\mathbb{U}/A$) for a small substructure A. A set *divides* if for some l it has an arbitrarily large set of l-wise disjoint conjugates (i.e., any l have empty intersection).

We will consider ideals of $\mathbb{U}$-definable sets (of some sort S). Say I is *definably generated* if it is generated by a definable family of definable sets. Say I is $\bigvee$-definable if it is generated by some bounded family of definably generated ideals.

Equivalently, for any formula definable $D \subset S \times S'$, $\{b \in S' : S(b) \in I\}$ is $\bigvee$-definable. If I is Aut($\mathbb{U}/A$)-invariant, then $\{b \in S' : S(b) \in I\}$ is in fact $\bigvee$-definable over A.

Dually, I determines a partial type over $\mathbb{U}$, generated by the complements of the definable sets in I. Any extension of this partial type is called *I-wide*. We say a/A is I-wide if a does not lie in any A-definable set lying in I. Note that tp(a/A) will then extend to an I-wide complete type over $\mathbb{U}$.

If $f : S \to S'$ is a 0-definable surjective map, and I is a $\bigvee$-definable ideal, let $f_*I = \{D : f^{-1}D \in I\}$. This is a $\bigvee$-definable ideal on S', proper if I is proper. If c/A is I'-wide, then $c = f(b)$ for some I-wide b/A.

If I, I' are two ideals (on S, S'), we can define an ideal $I \otimes I'$ on $S \times S'$, generated by the sets $D \subset S \times S'$ such that for some $D_1 \in I$, for all $a \in S \smallsetminus D_1$, $D(a) \in I'$. So if a/A is I-wide and $b/A(a)$ is I'-wide, then $(a, b)/A$ is $I \otimes I'$-wide. Conversely, if $(a, b)/A$ is $I \otimes I'$-wide, then a/A is I-wide, and — assuming I' is $\bigvee$-definable — $b/A(a)$ is I'-wide: to see the last statement, if $b \in D(a) \in I'$, then since I' is $\bigvee$-definable, there exists $\theta(x)$ true of a such that $D(a') \in I'$ for all $a' \in \theta$; let $D' = \{(a', b') : b' \in D(a'), a' \in \theta\}$; then $D' \in I \otimes I'$; and $(a, b) \in D'$.

Inductively, we define $I^{\otimes n}$, $I^{\otimes(n+1)} = I^{\otimes n} \otimes I$. We will say $b = (b_1, \ldots, b_n)$ is I-wide if it is $I^{\otimes n}$-wide.

Example A.7. Let us mention here some canonical ideals, relative to a given complete type p. There is Shelah's forking ideal I_{sh}, generated by the set $\mathrm{Div}(p)$ of formulas that divide (over $\varnothing$). Given any invariant measure μ (such that p is wide), we have the ideal I_μ of all formulas of μ-measure zero. If μ is definable, then I_μ is $\bigwedge$-definable. We have $\mathrm{Div}(p) \subseteq I_{\mathrm{sh}} \subseteq I_\mu$, for any invariant measure μ.

If I is an ideal on S', let $\mathrm{SDiv}(I)$ be the family of generically I-dividing subsets of S; i.e., the family of sets $Q(b)$, $b \in S'$, Q an A-definable subset of $S \times S'$, such that for some n, for any $I^{\otimes n}$-wide $(b_1, \ldots, b_n)$ with $\mathrm{tp}(b/A) = \mathrm{tp}(b_i/A)$, $\bigcap_{i=1}^n Q(b_i) = \varnothing$. Note that if $I \subset J$ then $I^{\otimes n} \subset J^{\otimes n}$, so $\mathrm{SDiv}(I) \subset \mathrm{SDiv}(J)$.

Let $\hat{I}$ be the ideal generated by $\mathrm{SDiv}(I)$. We have $\mathrm{SDiv}(I) \subseteq \mathrm{Div}$ and so $\hat{I} \subseteq I_{\mathrm{sh}}$. If I is $\bigvee$-definable over A, so are $\mathrm{SDiv}(I)$ and $\hat{I}$.

Definition A.8. Let $R \subset P \times P'$ be an invariant relation over A, and let I be a $\bigvee$-definable ideal on P. Say R holds *I-almost always* if for any $c \in P'$, for any $b \in P$ with $b/A(c)$ I-wide, we have $R(b, c)$. Say R holds *I-almost always in the strong sense* on $P \times P'$ if the transpose $R^t = \{(y, x) : (x, y) \in R\}$ holds $\hat{I}$-almost always.

Explicitly, R holds *I-almost always in the strong sense* on $P \times P'$ if whenever $(b, c) \in P \times P' \smallsetminus R$, there exists an A-definable local $Q \subset P \times P'$ and $n \in \mathbb{N}$ such that $(b, c) \in Q$, and for any $I^{\otimes n}$-wide n-tuple $(b_1, \ldots, b_n)$, $P \cap \bigcap_{i=1}^n Q(b_i) = \varnothing$.

If $R \subset S \times S'$ is an invariant relation, I a $\bigvee$-definable ideal on S, and $P \subseteq S$, $P' \subseteq S'$ invariant sets, we will also say that R holds *I-almost always in the strong sense on $P \times P'$* if $R \cap (P \times P')$ does.

Lemma A.9. *Assume R holds I-almost always in the strong sense on $S \times S'$.*

(1) *R holds I-almost always.*

(2) *If* $\mathrm{tp}(c/A(b))$ *does not divide over A, and* $\mathrm{tp}(b/A)$ *is I-wide, then $R(b, c)$.*

Proof. (1) Suppose not; let Q, n be as in Definition A.8. Let $b_1 = b$. Inductively find b_k such that $Q(b_k, c)$ and b_k is wide over $A(c, b_1, \dots, b_{k-1})$; this is possible since $Q(c)$ is wide. But then $c \in \bigcap_{i=1}^n R(b_i)$, a contradiction.

(2) Suppose $\neg R(b, c)$. Let Q be a definable set as in Definition A.8, so that for any I-wide $(b_1, \dots, b_n) \in S^n$, $\bigcap_{i=1}^n Q(b_i) = \varnothing$. As $\mathrm{tp}(b/A)$ is I-wide, one can find $b_i \models \mathrm{tp}(b/A)$ for $i \in \mathbb{N}$, such that $\mathrm{tp}(b_n/A(b_1, \dots, b_{n-1}))$ is wide. Then any subsequence of length n of this infinite sequence is $I^{\otimes n}$-wide, so the intersection of $Q(b_i)$ over any such subsequence is empty. It follows that $\mathrm{tp}(c/A(b))$ divides over A. $\square$

A.10. *Stable invariant local relations.*

Definition A.11. Two definable relations $P(x, y)$, $Q(x, y)$ are *stably separated* if there is no sequence of pairs (a_i, b_i), $i \in \mathbb{N}$, with $P(a_i, b_j)$ and $Q(a_j, b_i)$ for $i < j \in \mathbb{N}$.

Let $R \subset S \times S'$ be an $\mathrm{Aut}(\mathbb{U}/A)$-invariant relation.

Definition A.12. R is *stable* if whenever $(a, b) \in R$ and $(c, d) \in (S \times S') \smallsetminus R$, then there exist A-definable sets Q, Q' such that $Q(a, b)$, $Q'(c, d)$ and Q, Q' are stably separated.

Remark A.13. R is stable iff there is no indiscernible sequence (x_i, y_i) such that for $i \neq j$, $R(x_i, y_j)$ iff $i < j$.

Proof. If no such indiscernible sequence exists, then whenever $(a, b) \in R$ and $(c, d) \in (S \times S') \smallsetminus R$, $\mathrm{tp}(a, b)$ and $\mathrm{tp}(c, d)$ must be stably separated; by compactness, for some definable P approximating $\mathrm{tp}(a, b)$ and Q approximating $\mathrm{tp}(c, d)$, P, Q are stably separated. Conversely, if (a_i, b_i) is an indiscernible sequence as in the remark, then $\mathrm{tp}(a_1, b_2)$ is not stably separated from $\mathrm{tp}(a_2, b_1)$ though $R(a_2, b_1)$ and $\neg R(a_1, b_2)$. $\square$

Theorem A.14. *Let $\mathbb{U}$ be a local structure, with (♣). Let $\mathfrak{f}$ be a family of invariant stable local relations on $S \times S'$. Let $E_{\mathfrak{f}}$ be the intersection of all cobounded invariant local equivalence relations on S, such that each class is a Boolean combination of a bounded number of sets $R(b) \subset S$, $R \in \mathfrak{f}$. Then for each complete type $\boldsymbol{P}$ in S, there exists a proper, $\bigvee$-definable ideal $I(\boldsymbol{P})$ on S, satisfying*:

If $R \in \mathfrak{f}$, $P \subset \boldsymbol{P}$ is an $E_{\mathfrak{f}}$-class, and Q is an $E_{\mathfrak{f}^t}$-class on S', then either R holds almost always in the strong sense for $I(\boldsymbol{P})$ on $P \times Q$, or $\neg R$ does. $(*)$

Also, symmetry holds: if for P, Q as above, if $\bar{Q}$ is a complete type with $Q \subset \bar{Q}$, then on $P \times Q$, R holds almost always for $I(\boldsymbol{P})$ iff R^t holds almost always for $I^t(\bar{Q})$.

Remark A.15. $E_{\mathfrak{f}}$ has a distinguished class S^- such that for any $R \in \mathfrak{f}$, $\neg R$ holds almost always on $S^- \times S'$ in the strong sense for I_S. Away from this class, $E_{\mathfrak{f}}$ is a local relation. (See the proof above Lemma A.20.)

We remark that there also exists a canonical proper $\bigvee$-definable ideal I_S, such that the dichotomy $(*)$ and symmetry hold I_S-almost always. However it may trivialize certain types on S.

Though the proofs go through for any $\mathfrak{f}$, we will assume below that $\mathfrak{f} = \{R\}$ to simplify notation. (In fact the theorem reduces easily to the case that $\mathfrak{f}$ is finite; and then — replacing S by $S\times\mathfrak{f}$, and considering the relation $\hat{R}((x, R), y) \Longleftrightarrow R(x, y)$ — to the case that $\mathfrak{f}$ has a single element R.)

We will use the space $S_D(\mathbb{U})$ of all bounded global types on a sort D, i.e., types containing a formula implying $d(x, a) \le n$ for some a, n. If x is a variable of sort D, we will also write $S_x(\mathbb{U})$. Let $(d_p x)R = \{b : R(x, b) \in p\}$. If $(d_p x)R = (d_{p'} x)R$, we say p, p' define the same R-type. We do not define a topology on the set of global R-types.

Lemma A.16. *Let M be a countable model. Let $R'(x, y)$, $R(x, y)$ be definable relations* (*of which at least one is local*). *Assume $R'(x, y)$ and $R(x, y)$ are stably separated. Then for any type p over M there exists a finite Boolean combination Y of sets $R(x, c_i)$ with $c_i \in M$, such that $d_p y R' \Longrightarrow Y$ while Y, $d_p y R$ are disjoint.*

Proof. Let $c \models p|M$. Define $a_n, b_n, c_n \in M$ recursively. Given $c_1, \dots, c_n$, the equivalence relation: $\bigwedge_{i\le n} R(x, c_i) \Longleftrightarrow R(x', c_i)$ has at most 2^{2n} classes; if none of these classes meets both $d_p y R$ and $d_p y R'$, then some union Y of these classes contains $d_p y R$ and is disjoint from $d_p y R'$, and the lemma is proved. Otherwise, choose a_n, b_n such that $d_p y R(a_n)$, $d_p y R'(b_n)$, while a_n, b_n lie in the same sets $R(x, c_i)$, $i \le n$. Then, find c_{n+1} such that $R'(d, c_{n+1}) \Longleftrightarrow R'(d, c)$, where $d \in \{a_i, b_i : i \le n\}$.

For $n < k$ we have $R'(b_n, c_k)$. Applying Ramsey with respect to the question R and refining the sequence (a_n, b_n, c_n), we may assume that $R(b_n, c_k)$ for all $n > k$ or for no $n > k$; but the former is impossible since R', R are stably separated. So $\neg R(b_n, c_k)$ for all $n > k$.

Since a_n, b_n have the same R-type over the smaller c_i, it follows that $\neg R(a_n, c_k)$ for $n > k$. But for $n < k$ we have $R'(a_n, c_k)$; so the sequence (a_n, c_n) contradicts the stable separation of R', R. □

Corollary A.17. *Assume L is countable. Let R', R be stably separated local definable relations on $S \times S'$. There does not exist an uncountable set $W \subset S_x(M)$ such that for $p \ne p' \in W$, for some $b \in M$, $R'(x, b) \in p$ while $R(x, b) \in p'$.*

Proof. Let Y_p be an M-definable set such that $d_p R' \to Y_p \to \neg d_p R$ (Lemma A.16). There are only countably many choices for Y_p, so there will be $p, p' \in W$ with $Y_p = Y_{p'}$. Now if $R'(x, b) \in p$ then $b \in Y_p = Y_{p'}$ so $\neg R(x, b) \in p'$. □

It follows that there is no map f from the full binary tree $2^{<\omega}$ into S', such that for each branch $\eta \in 2^\omega$,

$$\bigwedge R'(x, f(\eta|n+1) : \eta(n) = 0) \wedge \bigwedge R(x, f(\eta|n+1) : \eta(n) = 1)$$

is consistent. By compactness, for some finite n, no such map exists for the height-n tree 2^n. We define the *rank* of a partial type W to be the maximum m such that there exists $f : 2^m \to S'$, with

$$W \wedge \bigwedge R'(x, f(\eta|n+1) : \eta(n) = 0) \wedge \bigwedge R(x, f(\eta|n+1) : \eta(n) = 1)$$

consistent for each $\eta \in 2^m$.

Let R be a stable invariant relation on $S \times S'$.

Lemma A.18. *Let p, q be types over $\mathbb{U}$. Assume that for any stably separated local definable ϕ, ψ, for some $e = e_{\phi,\psi}$ we have $e \subset p, q$ and $\mathrm{rk}_{\phi,\psi}(p) = \mathrm{rk}_{\phi,\psi}(e) = \mathrm{rk}_{\phi,\psi}(q)$. Then $p|R = q|R$.*

Proof. Let $c \models p$ and $d \models q$. Suppose $p|R \neq q|R$. Then for some $b \in \mathbb{U}$, $\mathrm{tp}(b, c)$ implies R but $\mathrm{tp}(b, d)$ implies $\neg R$. As R is stable, $\mathrm{tp}(b, c)$ and $\mathrm{tp}(b, d)$ are stably separated; hence by compactness, some $\phi(x, y) \in \mathrm{tp}(b, c)$ and $\psi(x, y) \in \mathrm{tp}(b, d)$ are stably separated. Let $e = e_{\phi,\psi}$, $l = \mathrm{rk}_{\phi,\psi}(e)$. Let $[\phi(x, b)]$ be the set of types extending $\phi(x, b)$. It follows that either $\mathrm{rk}_{\phi,\psi}(e \cap [\phi(b, x)]) < l$ or $\mathrm{rk}_{\phi,\psi}(e \cap [\psi(b, x)]) < l$. But $\mathrm{rk}_{\phi,\psi}(p) = \mathrm{rk}_{\phi,\psi}(q) = l$, a contradiction. □

Remark (uniqueness of finitely satisfiable extensions). Thus if e is a partial type, $e \subseteq p, q$, and $\mathrm{rk}_{\phi,\psi}(p) = \mathrm{rk}_{\phi,\psi}(e) = \mathrm{rk}_{\phi,\psi}(q)$ for all stably separated (ϕ, ψ), then for all stable invariant relations R we have $p|R = q|R$. This hypothesis holds if e is a type over a model M, and p, q extend e and are finitely satisfiable in M.

Remark (determination by Ind-definable part). We can also deduce that for any global p, p', if p' contains all schemes

$$\{\psi(x, b) : \theta(b)\}$$

that are contained in p, then for any stable invariant relations R, we have $p|R = p'|R$. For this, for each stably separated pair (ϕ, ψ), we look at the deepest (ϕ, ψ)-binary tree *contained* in p (rather than consistent with p).

For any partial type Q, we let $\hat{Q}$ denote the set of types over $\mathbb{U}$ extending Q.

Proposition A.19. *Let R be a stable local invariant relation on $S \times S'$. Assume (♣). Let X be a nonempty closed invariant subset of $\hat{S}$. Let $X|R = \{(d_p x)R : p \in X\}$.*

Then $1 \Rightarrow 2 \Rightarrow 3$:

(1) *X is minimal.*

(2) *for any stably separated ϕ, ψ defined over A, $\mathrm{rk}_{\phi,\psi}(p)$ is constant (does not depend on $p \in X$).*

(3) *$X|R$ has cardinality bounded independently of $\mathbb{U}$; in fact $|X|R| \leq \aleph_0^{|L|}$.*

Moreover, a minimal nonempty closed invariant subset of X exists.

Proof. (1) implies (2) since the set of elements of X of (ϕ, ψ)-rank $\geq n$ is a closed, invariant subset of X.

Now assume (2). Fix ϕ, ψ stably separated, and say $\mathrm{rk}_{\phi,\psi}(p) = m$ for $p \in X$. For each ball B of the metric d, the intersection of B, X and the complement of all definable sets of (ϕ, ψ)-rank $\leq m$ is empty; by (local) compactness, $B \cap X$ is covered by finitely many definable sets of (ϕ, ψ)-rank $\leq m$. Thus X is covered by countably many such definable sets, say $e(\phi, \psi, l)$, $l \in \mathbb{N}$. Each p now determines a function $\chi_p : (\phi, \psi) \mapsto l$, where l is least such that $p \in e(\phi, \psi, l)$. But in turn $p|R$ is determined by this function. For if $p, p' \in X$ and $\chi_p = \chi_{p'}$, then by Lemma A.18, $p|R = p'|R$. This proves (3).

For the moreover, given a complete type P, let $\hat{P}$ be the set of types over $\mathbb{U}$ compatible with P. Then X meets some $\hat{P}$ nontrivially, so letting $Z = X \cap \hat{P}$ it suffices to show that any nonempty closed invariant subset Z of $\hat{P}$ contains a minimal nonempty closed invariant subset. Fix $b \in P$, and let B be the ball defined by $d(x, b) \leq 2m_0$. By (♣), any type p over $\mathbb{U}$ meets some m_0-ball; by saturation of $\mathbb{U}$, this m_0-ball contains a $P(\mathbb{U})$-point a; so $d(x, a) \leq 2m_0$ is compatible with p. By invariance, $d(x, b) \leq 2m_0$ is compatible with some $p' \in Z$. Thus $\hat{B} \cap Z \neq \varnothing$ (where $\hat{B}$ is the set of all types over $\mathbb{U}$ of elements of B). So if Z_i is a descending chain of nonempty closed invariant subsets of $S^P_R(\mathbb{U})$, then $Z_i \cap \hat{B}$ is nonempty, and as $\hat{B}$ is compact, $\bigcap Z_i \cap \hat{B}$ is nonempty, and in particular $\bigcap Z_i$ is nonempty. Thus by Zorn's lemma a minimal element exists. □

Let S, S' be sorts, and $R \subset S \times S'$ be invariant, stable. Let Gen^R be the set of all restrictions $p|R$, where p is a global type of S and $p|R$ has a small orbit under $\mathrm{Aut}(\mathbb{U})$. (The total number of orbits is small, say by Lemma A.18, so Gen^R is small.) When relativizing to a small set A, so R is $\mathrm{Aut}(\mathbb{U}/A)$-invariant, we write Gen^R_A.

Any type P on S extends to some element of Gen^R, by Proposition A.19. It follows that for any $\equiv_{\mathrm{Las}}$-class X on S there exists an element q_X of Gen^R such that for any small N, $q_X|N$ is realized in X. Indeed *some* $\equiv_{\mathrm{Las}}$-class of P has this property; since all $\equiv_{\mathrm{Las}}$-classes in P are conjugate, all have it.

Similarly define ${}^R\mathrm{Gen} = \mathrm{Gen}^{R^t}$ on S'.

Define an equivalence relation $E_{\mathfrak{f}}$ on S by: $(a, b) \in E_{\mathfrak{f}}$ iff for all $p \in {}^R\mathrm{Gen}$ and $R \in \mathfrak{f}$, $(d_p y)R(a, y) \iff (d_p y)(R_b, y)$; and dually define E_{R^t} on S'. $E_{\mathfrak{f}}$ is cobounded since ${}^R\mathrm{Gen}$ is bounded. $E_{\mathfrak{f}}$ is local since R is local: if $aE_{\mathfrak{f}}b$ then for some c, $R(a, c)$ and $R(b, c)$, so $d(a, b) \leq d(a, c) + d(b, c)$.

We say that $q|R$ is consistent with an invariant set Z if any small subset q_0 of $q|R$ is realized by some element of Z.

Lemma A.20 (symmetry and uniqueness). *Any $E_{\mathfrak{f}}$-class on S is consistent with a unique $q \in \mathrm{Gen}^R$. If $q \in \mathrm{Gen}^R$, $q' \in {}^R\mathrm{Gen}$, $a \in S$, $a' \in S'$, and q is consistent with $E_{\mathfrak{f}}(a)$, and q' with $E_{\mathfrak{f}^t}(a')$, then $d_{q'}yR(a, y) \iff d_q xR(x, b)$.*

Proof. We prove the symmetry statement first, following the standard route. Suppose for contradiction that it fails for q, q', a, a'. Say $d_{q'} y R(a, y)$ holds but $dq_x R(x, b)$ fails. Construct a_n, a'_n so that $a_n \models q|A(a'_i : i < n)$, $a_n E_{\mathfrak{f}} a$, and $a'_n \models q'|A(a_i : i < n)$, $a'_n E_{R^t} a'$. Then since $a_n E_{\mathfrak{f}} a$, $d_{q'} y R(a_n, y)$ holds, and similarly $d_q x R(x, a'_n)$ fails. Thus if $i > n$ then $R(a_n, a'_i)$ holds but $R(a_i, a'_n)$ fails. This contradicts the stability of R.

We have already shown that there exists $q' \in {}^R\mathrm{Gen}$ consistent with the Lascar type $E_{\mathfrak{f}^t}(a')$. Now if $q_1, q_2 \in \mathrm{Gen}^R$ are both consistent with $E_{\mathfrak{f}}(a)$, then by symmetry we have $d_{q_1} x R(x, b) \iff d_{q'} y R(a, y) \iff d_{q_2} x R(x, b)$. Thus $q_1 = q_2$. □

Because of this lemma, if χ is an $E_{\mathfrak{f}}$-class and q is the unique element of Gen^R consistent with it, we can write $(d_\chi x) R(x, y)$ for $(d_q x) R(x, y)$.

Let χ be an $E_{\mathfrak{f}}$-class, consistent with q. Let M be a substructure such that for any two elements $q_1 \neq q_2 \in \mathrm{Gen}^R$, there exists $b \in M$ with $R(x, b) \in q_1$ but $R(x, b) \notin q_2$, or vice versa. Let $E^{\mathfrak{f}}_M$ be the equivalence relation: $a E^{\mathfrak{f}}_M b$ iff for any $R \in \mathfrak{f}$ and $b \in M$, $R(a, b) \iff R(a, b')$. Then χ is a cobounded equivalence relation, each class is a bounded Boolean combination of sets $R^t(b)$, and $E^{\mathfrak{f}}_M$ refines $E_{\mathfrak{f}}$. Indeed by construction a unique element $q \in \mathrm{Gen}^R$ will be consistent with a given $E^{\mathfrak{f}}_M$-class χ. So for any $q' \in {}^R\mathrm{Gen}$, let d be such that $\mathrm{tp}(d/M)$ is consistent with q'; then for $a \in \chi$, $R(a, y) \in q'$ iff $R(x, d) \in q$.

Since all $E_{\mathfrak{f}}$ classes of a complete type P over A are $\mathrm{Aut}(\mathbb{U}/A)$-conjugate, it follows from uniqueness that all elements q of Gen^R consistent with P are $\mathrm{Aut}(\mathbb{U}/A)$-conjugate.

We *choose* a minimal nonempty closed $\mathrm{Aut}(\mathbb{U}/A)$-invariant set $X = X_P$ of global types extending P, as in Proposition A.19. By this lemma, for any ϕ, ψ, $\beta_p(\phi, \psi) = \mathrm{rk}_{\phi,\psi}(p)$ does not depend on the choice of $p \in X$. Let $I(P) = I(X_P)$ be the ideal generated by all definable sets D such that for some ϕ, ψ we have $\mathrm{rk}_{\phi,\psi}(D) < \beta_p(\phi, \psi)$.

Lemma A.21 (dividing). *Let q' be a global type of elements of S'. Assume that $q'|R^t \in {}^R\mathrm{Gen}$, P is an $E_{\mathfrak{f}}$-class, and $R(a, y) \in q'$ for $a \in P(\mathbb{U})$. For $i \in \omega_1$, let $b_i \models q'|A(b_j : j < i)$. Then for any $a \in P(\mathbb{U})$, for cofinally many $\alpha \in \omega_1$ we have $R(a, b_\alpha)$.*

Proof. Redefine b_i (without changing the type of the sequence) as follows: let $M_i \prec \mathbb{U}$ be a small model containing a_j for $j < i$, and let $b_i \models q'|M_i$. Let $M = \bigcup_{i<\omega_1} M_i$. For any pair (ϕ, ψ), for some $i < \omega_1$, we have $\mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M_i)) = \mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M))$. Since ω_1 has uncountable cofinality, for some $\alpha < \omega_1$, for any ϕ, ψ, $\mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M_\alpha)) = \mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M))$. Since $M_\alpha \prec \mathbb{U}$, there exists a global type q extending $\mathrm{tp}(a/M_\alpha)$ such that $\mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M_\alpha)) = \mathrm{rk}_{\phi,\psi}(q)$. By Lemma A.18, $q|R$ is uniquely determined. On the other hand since $q'|R^t \in \mathrm{Gen}^R_A(S')$, it is clear that $q'|R^t \in \mathrm{Gen}^R_M(S')$. Since $R(a, y) \in q'$, by Lemma A.20, $R(x, b) \in q$ if $\mathrm{tp}(b/M_\alpha)$

is consistent with q'. Hence $R(x, b_i) \in q$ for $i \geq \alpha$. But we can also construct a global type q^+ extending $\mathrm{tp}(a/M_{\alpha+1})$ with $\mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M_{\alpha+1})) = \mathrm{rk}_{\phi,\psi}(q^+)$. As $\mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M_{\alpha+1})) = \mathrm{rk}_{\phi,\psi}(\mathrm{tp}(a/M_\alpha))$, it follows that $q = q^+$; as $R(x, b_\alpha) \in q$ we have $R(x, b_\alpha) \in q^+$, i.e., $R(a, b_\alpha)$. $\square$

It follows from Lemma A.21 (as well as from Lemma A.20, as we saw before) that $(d_p y)R(x, y)$ is a bounded (but infinitary) Boolean combination of instances of $R(x, b)$; namely $(d_p y)R(a, y)$ iff $R(a, b_j)$ holds for cofinally many j, where (b_j) is a sufficiently long sequence as in the lemma.

Proof of Theorem A.14. We will use the equivalence relation $E_{\mathfrak{f}}$ and the ideals $I(P)$ defined above Lemma A.21. We have to show:

> If $R \in \mathfrak{f}$, $P \subset \boldsymbol{P}$ is an $E_{\mathfrak{f}}$-class, and Q is an $E_{\mathfrak{f}^t}$-class on S', then either R holds almost always in the strong sense for $I(P)$ on $P \times Q$, or $\neg R$ does. $(*)$

Pick $p \in X(\boldsymbol{P})$, and $p' \in X(\bar{Q})$ (with respect to ${}^tR$). By definition of $E_{\mathfrak{f}}$, for any $a \in P$, $p'(y)$ implies $R(a, y)$, or else for any $a \in P$, $p'(y)$ implies $R(a, y)$. Without loss of generality the latter holds. Now suppose $\neg R(c, b)$ holds with $c \in P, b \in Q$. As $p'(y)$ implies $R(a, y)$ and $E_{\mathfrak{f}^t}(a, c)$, $p'(y)$ also implies $R(c, y)$. Let $r = \mathrm{tp}(c, b/A)$. We have to show that the condition in Definition A.8 holds, i.e., that for some n, and some $D \in r$, $\bigcup D(x, y_j) \cup \neg I_{\mathfrak{f}^t}^{\otimes n}(y_1, \ldots, y_n)$ is inconsistent. Otherwise, there exists a sequence $c, b_1, b_2, \ldots$ with $b_k/A(b_1, \ldots, b_{k-1})$ wide for $I_{\mathfrak{f}^t}$ for each k, and $r(c, b_i)$ holds for each i. Let σ be an automorphism taking (c, b) to (c, b_1). Then $q' = \sigma(p')$ is a global type, $q'|R^t \in \mathrm{Gen}$, consistent with $E_{\mathfrak{f}^t}$-class of $\sigma(b_1)$, and $q'(y)$ implies $R(c, y)$ (since $\sigma(c) = c$). By Lemma A.21, $R(c, b_i)$ holds for some i. But r is a complete type, and cannot be consistent with both $\neg R(c, b)$ and $R(c, b_i)$. This shows that $\bigcup D(x, y_j) \cup \neg I_{\mathfrak{f}^t}^{\otimes n}(y_1, \ldots, y_n)$ is indeed inconsistent.

We saw that $(d_p y)R(x, y)$ is a bounded Boolean combination of instances of $R(x, b)$; hence any $E_{\mathfrak{f}}$-class can be expressed as Boolean combination of a bounded number of sets $R(b) \subset S$, $R \in \mathfrak{f}$. Given this, the finest cobounded equivalence relation with this property refines $E_{\mathfrak{f}}$, and so also satisfies $(*)$. $\square$

Remark A.22. Let $p(x, y)$ be a partial type. Then there exists a unique smallest stable invariant relation P containing p. (I.e., p implies P.) P is F_σ. Likewise for "equational" in place of stable.

Proof. We prove the stable case; the equational case is the same, with $a_0 = a$, $b_0 = b$ below. For any invariant relation $P(x, y)$, let $P'(a, b)$ hold iff there exists an indiscernible sequence of pairs (a_i, b_i) with $a_1 = a$, $b_0 = b$, and $P(a_0, b_1)$. Clearly P' is $\bigwedge$-definable if P is; and P is stable iff $P = P'$. Also if $P = \bigvee_j P_j$ then $P' = \bigvee_j P_j'$; and the operation $P \mapsto P'$ is monotone. So let $P_0 = p$, $P_{n+1} = P_n'$

and $P = \bigcup_{n\in\mathbb{N}} P_n$. Then P is F_σ and stable, and contained in any stable invariant relation containing p. $\square$

Presumably P is usually not $\bigwedge$-definable (for instance when p implies $\equiv_{\mathrm{Las}}$ and $\equiv_{\mathrm{Las}}$ is not $\bigwedge$-definable). Note that $\equiv_{\mathrm{Las}}$ is itself a stable invariant relation.

A.23. ***$\bigwedge$-definable stable relations.*** We will discuss a stable, $\bigwedge$-definable relation $R(x, y)$; the results go through in the same way for a set $\mathfrak{f}$ of such relations. We assume for simplicity that the language L and base set A are countable, so $R = \bigcap_n R_n$ for some sequence R_n of definable relations, with $R_1 \supset R_2 \supset \cdots$; the general case reduces immediately to this. We work with a universal domain $\mathbb{U}$.

First we note that the p-definition of R is $\bigwedge$-definable, for any type p.

Lemma A.24 (definability). *Let $p \in S_x(M)$. Let $R = \bigcap_n R_n$, with R_n definable.*

(1) *$d_p R$ is $\bigwedge$-definable over M; it is an intersection of Boolean combinations of sets $R_n(c)$ with $c \in M$.*

(2) *In fact for any $m \in \mathbb{N}$ there exists $n = n(m)$ and a finite Boolean combination Y of sets $R_n(x, c_i)$, $c_i \in M$, such that $d_p R \to Y \to d_p R_m$.*

Proof. It suffices to prove (2). By stability, there is no sequence d_n, e_k with $\neg R_m(d_n, e_k)$ for $k > n$ and $R(d_n, e_k)$ for $k < n$. By compactness, for some n_0, there is no sequence with $\neg R_m(d_n, e_k)$ for $k > n$ and $R_{n_0}(d_n, e_k)$ for $k < n < n_0$. Thus $\neg R_m$, R_{n_0} are stably separated. By Lemma A.16, there exists a finite Boolean combination Y of sets $R_{n_0}(x, c_i)$, $c_i \in M$, such that $d_p R_{n_0} \to Y \to d_p R_m$. $\square$

Lemma A.25. *Any $E_\mathfrak{f}$-class of elements of P is $\bigwedge$-definable with parameters, on any complete type P. It is cut out by certain sets of the form $(d_q y)R(x, y)$.*

Proof. Let P be a complete type of S.

We can find $a \in P$ such that $Q(a) = \{q \in {}^R\mathrm{Gen} : a \in (d_q y)R(x, y)\}$ is maximal, i.e., not properly contained in any $Q(a')$ (with $a' \in P$). This uses Zorn's lemma, and the fact that $(d_q y)R(x, y)$ is $\bigwedge$-definable, so if $(d_q y)R(a_i, y)$ and $\mathrm{tp}(a_i/M)$ approaches $\mathrm{tp}(a/M)$ in the space of types over M, then $(d_q y)R(a, y)$.

Let $Q = Q(a)$. Now $aE_\mathfrak{f}b$ iff for each $q \in Q$, $(d_q y)R(b, y)$. So the $E_\mathfrak{f}$-class of a is $\bigwedge$-definable.

Since all $E_\mathfrak{f}$-classes in P are conjugate, all $E_\mathfrak{f}$-classes in P are $\bigwedge$-definable. As P was arbitrary, the lemma follows. $\square$

Corollary A.26. *If $a \equiv_{\mathrm{lc}} b$ then $(a, b) \in E_\mathfrak{f}$.*

Proof. In any case $a \equiv_{\mathrm{lc}} b$ implies that a, b have the same complete type; so it suffices to show this for $a, b \in P$, where P is a complete type.

Define aEb iff $\mathrm{tp}(a/c) = \mathrm{tp}(b/c)$ for any $E_\mathfrak{f}$-class c (i.e., there exists an automorphism fixing c and taking a to b). Clearly $E \subset E_\mathfrak{f}$. Let $\{C_i : i \in I\}$ list all the classes. Then aEb iff for each i, $(\exists c)(\exists d)(c, d \in C_i \wedge ac \equiv bd)$. Since each C_i

is $\bigwedge$-definable by Lemma A.25, E is $\bigwedge$-definable. Since the number of classes C_i is bounded, and elements with the same type over some representative $c_i \in C_i$ also have the same type over C_i, it is clear that E is cobounded. Hence $\equiv_{\mathrm{lc}} \subset E$, so $\equiv_{\mathrm{lc}} \subset E_{\mathfrak{f}}$. $\square$

From this and Theorem A.14 we obtain:

Theorem A.27 (locally compact equivalence relation theorem). *Let $\mathfrak{f}$ be a nonempty family of $\bigwedge$-definable stable local relations on $S \times S'$. Assume S' is a complete type. There exists a proper $\bigvee$-definable ideal I' of definable subsets of S', such that if $R \in \mathfrak{f}$, and P, Q are classes of $\equiv_{\mathrm{lc}}$ on S, S' respectively, then R holds almost always on $P \times Q$ in the strong sense for I', or $\neg R$ does. Symmetry holds as in Theorem A.14. Also, the analogue of Remark A.15 is valid.*

In particular, fix a and assume $\mathrm{tp}(a/A)$ forms a single $\equiv_{\mathrm{lc}}$-class; then for b such that $\mathrm{tp}(a/Ab)$ or $\mathrm{tp}(b/Aa)$ does not divide over A, the truth value of $R(a, b)$ depends only on $\mathrm{tp}(b)$.

Note that in the case of a definable measure, the measure 0 ideal is $\bigwedge$-definable and so in general properly contains the ideal I' we found here; they coincide only when both are definable.

Corollary A.28. *Let $R = \bigcap_n R_n$ be a $\bigwedge$-definable stable local relations on $S \times S'$. Assume S' is a complete type. Let P, Q be classes of $\equiv_{\mathrm{lc}}$ on S, S', and assume R holds almost always on $P \times Q$, as in Theorem A.27. Then for each n there exists a neighborhood U of (P, Q) such that if $(P', Q') \in U$ then R_n holds almost always on $P' \times Q'$.*

Appendix B. Over a model

The entire thrust of this paper is to give a *lightface* account of higher measure amalgamation, choosing no constants.

Here we record the much better understood situation *over a model* in a similar language. The idea is not to study the correlations in detail, but simply to take an elementary submodel M_0 as if it were completely known, and describe the situation almost everywhere “above M_0”, relying on the fact that anything that may happen with positive probability has already happened in M_0.

Theorem B.11 is a model-theoretic version of the hypergraph Szemerédi (or quasirandomness) lemma. The methods are essentially those of Theorem 5 of Towsner [73], and the results of Tao cited there. The results are valid only over a model, and in addition, only “almost everywhere”; they are blind to phenomena occurring on measure zero sets of n-types, and so cannot give a meaningful stationarity lemma valid for all types (or even for almost all n-tuples of 1-types, as opposed to almost all n-types).

We assume here that L is a countable language, T a complete theory, X, Y definable sets carrying definable measures μ_X, μ_Y. Form the multiple integral measures, and assume Fubini holds for the product measures on $X \times Y^n$, for each n. Let $\phi(x, y)$ be a definable relation on $X \times Y$.

Lemma B.1. *Let M be a countable model, and $\phi(x, y)$ a formula. Let*

$$B(\phi) = \left\{ \mathrm{tp}(a/M) : \mu_Y \phi(a, y) > 0, \bigwedge_{m \in M} \neg\phi(a, m) \right\}.$$

Then $B(\phi)$ has measure zero.

Proof. Note that $B(\phi)$ is a Borel subset of $S_x(M)$, in fact the intersection of an open set with a closed set. Fix ϕ, and let

$$B_\epsilon = \left\{ \mathrm{tp}(a/M) : \mu_Y \phi(a, y) \geq \epsilon, \bigwedge_{m \in M} \neg\phi(a, m) \right\}.$$

So as ϵ descends to 0, $B(\phi)$ is the increasing union of the sets B_ϵ, and it suffices to show that each closed set B_ϵ has measure zero, or just that $\mu_X(B_\epsilon) < \epsilon$. Fix $\epsilon > 0$, and let

$$X_\epsilon = \{\mathrm{tp}(a/M) : \mu_Y \phi(x, y) \geq \epsilon\}.$$

Let $n = n(\epsilon)$ be large, so that $\mu_X(X_\epsilon)(1-\epsilon)^n < \epsilon$, and set

$$W = \left\{ (x, y_1, \ldots, y_n) \in X \times Y^n : \bigwedge_{i=1}^{n} \neg\phi(x, y_i) \right\}.$$

Let $\mu = \mu_X \otimes \mu_Y \otimes \cdots \otimes \mu_Y$. Clearly, $\mu(W) \leq \mu_X(X)(1-\epsilon)^n < \epsilon$. Let

$$Y_\epsilon = \{y \in Y^n : \mu_X\{x : (x, y) \in W\} \geq \epsilon\}.$$

By Fubini, Y_ϵ cannot have full measure in Y^n.

So $Y' := Y \setminus Y_\epsilon$ is not a null set. Since μ_X is a definable measure and M is a model, we have $Y'(M) \neq \varnothing$. Thus for some $m_1, \ldots, m_n \in M$,

$$\mu_X\{x : (x, m_1, \ldots, m_n) \in W\} < \epsilon.$$

But $B_\epsilon \subset \{x : (x, m_1, \ldots, m_n) \in W\}$; so $\mu_X(B_\epsilon) < \epsilon$. Letting $\epsilon \to 0$ we see that $\mu_X(B(\phi)) = 0$. $\square$

Corollary B.2. *For almost all types* $\mathrm{tp}(a/M)$ *in X, any weakly random type (in Y) over Ma is finitely satisfiable in M.*

Proof. By Lemma B.1, $B := \bigcup_\phi B(\phi)$ has measure zero (here ϕ ranges over all formulas $\phi(x, y)$ over M). Assume $\mathrm{tp}(a/M) \notin B$. Let $\mathrm{tp}(b/Ma)$ be weakly random. Then for any formula $\phi(x, y) \in \mathrm{tp}(a, b/M)$, we have $\mu_Y \phi(a, y) > 0$, by weak randomness. Hence by definition of B, $\phi(a, m)$ holds for some $m \in M$. $\square$

Remark B.3. An independent family (E_a) of finite equivalence relations may, in general, be definable; then the effect of E_a cannot be accounted for before one is aware of the parameter a, and one cannot expect 3-amalgamation to hold over $M \cup \{a\}$, but only at best over $M \cup \mathrm{bdd}(a)$. Thus 4-amalgamation cannot hold over M, in general, if we attempt to amalgamate extensions that are not algebraically closed.

Lemma B.1 shows nevertheless that for *almost all types*, amalgamation *is* possible; for a realizing a random type over M, the finitely many classes of E_a will already be represented in M.

It will be useful to state a (tautological) measure-theoretic lemma on compatibility of conditional expectation with random fibers.

Lemma B.4. *Let $X \to Y \to Z$ be Borel maps between standard Borel spaces, and let μ_X be a Borel probability on X, with pushforwards μ_Y on Y and μ_Z on Z. For $z \in Z$, let Y_z be the fiber above z and let X_z be the fiber above the composed map $X \to Z$. Assume μ "disintegrates" as an integral over Z of a Borel family of measures μ_z on the fibers X_z (so $\mu_X = \int_Z \mu_z$). Let $\phi : X \to \mathbb{R}$ be a bounded Borel function, with expectation $E(\phi)$ on Y. For an L^1-function ψ on X_z, let $E_z(\psi)$ denote the expectation on Y_z with respect to ν_z. Let $\mu_{Y,z}$ be the pushforward of μ_z to Y_z. Then $\mu_Y = \int_{z\in Z} \mu_{Y,z}$; and for μ_Z-almost all points $z \in Z$, we have*

$$E(\phi)|Y_z =_{Y_z\text{-a.e.}} E_z(\phi|X_z).$$

Proof. When $Z = \{0, 1\}$, $Y = Y_0 \dot\cup Y_1$, and (if $\mu_Z(Y_0) > 0$, and pulling back Y_0 to $X_0 \subset X$) the statement is that $E(\phi)|Y_0 = E(\phi|X_0)$, which is clear. The general case follows by approximation.[9] □

B.5. Let $\mathcal{L}$ be a continuous logic language, and $\mathcal{T}$ be a stable theory of $\mathcal{L}$. We assume $\mathcal{T}$ eliminates quantifiers and imaginaries.

Assume given further a piecewise, partial interpretation $\mathcal{S}$ of $\mathcal{T}$ in T, namely a family $\mathcal{F}$ of maps from sorts of L to various sorts of $\mathcal{T}$, such that

(1) If $f : X \to Y$ and $g : X' \to Y'$ are in $\mathcal{F}$ then so is $f \times g : X \times X' \to Y \times Y'$.

(2) If $f : X \to Y$ lies in $\mathcal{F}$, and $g : Y \to Y'$ is a $(\bigwedge)$-definable map of $\mathcal{T}$, then $g \circ f$ lies in $\mathcal{F}$.

(3) The pullback of any $\mathcal{L}$-$\bigwedge$-definable subset of $\mathcal{T}$ under any $f \in \mathcal{F}$ is L-$\bigwedge$-definable.

[9] Or in Radon-Nikodym style: using separability of L^2, or countable generation of the algebra, it suffices to show for a Borel function ψ on Y that $\int_{Y_z} \psi(y)E(\phi) = \int_{Y_z} \psi(y)E_z(\phi|X_z)$, for almost all z. This in turn is equivalent to showing for any bounded Borel θ on Z that $\int_Z \theta(z) \int_{Y_z} \psi(y)E(\phi) = \int_Z \theta(z) \int_{Y_z} \psi(y)E_z(\phi|X_z)$. Now $E(\theta(z)\psi(y)\phi) = \theta(z)\psi(y)E(\phi)$, so the left-hand side is just $\int_X \theta(z)\psi(y)\phi(x)$. Similarly $E_z(\phi|X_z)\theta(z)\psi(y) = E_z(\phi|X_z)\theta(z)\psi(y))$, so the right-hand side is $\int_{z\in Z} \int_{X_z} \phi|X_z)\theta(z)\psi(y) = \int_X \theta(z)\psi(y)\phi(x)$ too.

By *partial* we mean that the maps f need not cover between them a full model of $\mathcal{T}$, but perhaps only a substructure.

When $N \models T$ is sufficiently saturated and homogeneous, and $A \subset N$ is countable, we have $\mathrm{dcl}(A) = \mathrm{Fix}\,\mathrm{Aut}(N/A)$. We denote by $\mathcal{S}(A)$ the definable closure of A within $\mathcal{S}$:

$$\mathcal{S}(A) = \{f(a) : f \in \mathcal{F}, a \in \mathrm{dcl}(A)\}.$$

Also write $\mathcal{S}_M(A) := \mathcal{S}(M \cup A)$.

When given a tuple $(a_1, \dots, a_n)$, write

$$a[n] := (a_1, \dots, a_n), \qquad a[n-i] := (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n).$$

We will use the fact that in a stable theory if a forks with c, then some formula $\phi(x, c) \in \mathrm{tp}(a/c)$ causes forking, i.e., $\phi(a, c)$ takes value > 0, and for any c', if $\phi(a, c')$ takes value > 0 then a, c' are not independent. (The forking is due to some $\psi(x, y)$; let $\phi_1(x, y, e)$ (with e a parameter in $\mathrm{bdd}(0)$) be the absolute value of the difference between $\psi(x, y)$ and the $\mathrm{tp}(a/\mathrm{bdd}(0))$-definition of ψ; then quantify out e, taking an appropriate supremum.)

Lemma B.6. *Let $M \subset N \models T$, $(a_1, \dots, a_n, c) \in N$, and assume $\mathrm{tp}(c/M(a_1, \dots, a_n))$ is finitely satisfiable in M. Then $\mathcal{S}_M(a[n])$ is independent from $\bigcup_i \mathcal{S}_M(a[n-i], c)$ over $\bigcup_i \mathcal{S}_M(a[n-i])$.*

Proof. Let $b \in \mathcal{S}_M(a[n])$, and let $\phi(y, u)$ be a formula of $\mathcal{L}$, where $u = (u_1, \dots, u_n)$. Also let $d_i \in \mathcal{S}_M(a[n-i], c)$, suppose $\phi(b, d_1, \dots, d_n) > 0$, and that $\phi(b, u) > 0$ causes forking over $\bigcup_i \mathcal{S}_M(a[n-i])$. We may write $d_i = f_i(a[n-i], c)$, where f_i is a $\bigwedge$-definable function. Since $\mathrm{tp}(c/M(a_1, \dots, a_n))$ is finitely satisfiable in M, there exists $c' \in M$ with $\phi(b, (f_i(a[n-i], c'))_i) > 0$; let $d'_i = f_i(a[n-i], c')$). Then $\mathrm{tp}_\phi(b/d'_1, \dots, d'_n)$ forks over $\bigcup_i \mathcal{S}_M(a[n-i])$. But this is a contradiction since $d'_i \in \mathcal{S}_M(a[n-i])$. $\square$

Compare [73, Lemma 4]. Note that Fubini is not required here.

Definition B.7. Let Q_k be a collection of k-types, closed under restrictions and permutations of variables, and set $Q = \bigcup_n Q_k$. Consider a downward-closed family S of subsets of $\{y_1, \dots, y_N\}$, all containing some base set s_0 and with $|s \smallsetminus s_0| \leq k$ for $s \in S$.

We say that Q is a $(\leq k, \infty)$-amalgamation family if for any such family S, and any map $j : S \to Q$ compatible with restrictions, such that $j(u) \in Q_{|u|}$, the union $\bigcup_{u \in S} j(u)$ is consistent, and in fact extends to an element of Q_N.

(This is equivalent to l-amalgamation for each $l \leq k+1$, over a base set, in the sense of [42].)

An n-tuple whose type is in Q_n will be called Q-independent.

Note that for $k > 2$, the hypothesis is incompatible with the presence of a linear ordering. However it holds in many simple theories, for instance pseudofinite fields over a base A such that definable and algebraic closure coincide over A. (Such a base exists for "most" but not all completions of the theory of pseudofinite fields.) A more refined version taking algebraic closure of each node in the system is valid in all pseudofinite fields, but we restrict ourselves to the simpler case.

Theorem B.8 (stationarity). *Let μ be a definable measure on a sort X. Let Q be a $(\leq n-1, \infty)$ -amalgamation family. Let $\phi_j(y_1, \ldots, y_n, x)$ be formulas with y_j a dummy variable not mentioned in ϕ_j, and let* $\mathrm{tp}(a^1, \ldots, a^n) \in Q_n$. *Then the quantity*

$$\mu\Big(\bigwedge_{j \leq n} \phi_j(a, x)\Big)$$

depends only on the n-tuple of $n-1$-subtypes of $\mathrm{tp}(a^1, \ldots, a^n)$ *and not on the full n-type.*

Proof. Write $p =_{<n} p'$ to mean that the two n-types agree on any restriction to $< n$ of the variables.

For the sake of readability we take $n = 3$ as a representative case, and write (a, b, c) for (a^1, a^2, a^3). Suppose (a, b, c) and (a', b', c') are Q-independent elements, and $\mathrm{tp}(a, b, c) =_{<3} \mathrm{tp}(a', b', c')$. We have to prove that for formulas ϕ_j as above, $\mu\big(\bigwedge_j \phi_j(a, b, c, x)\big) = \mu\big(\bigwedge_j \phi_j(a', b', c; x)\big)$.

We may assume here that $\mathrm{tp}(a, b, c, a', b', c')$ is Q-independent (using amalgamation over $\varnothing$).

We will construct $M = \{a_1, a_2, \ldots\}$ such that

(i) $\mathrm{tp}(a/M, b, c)$ is finitely satisfiable in M,

(ii) likewise for $\mathrm{tp}(a'/Mb'c')$, and

(iii) $\mathrm{tp}(abc/M) =_{<3} \mathrm{tp}(a'b'c'/M)$.

During the induction, at stage n, we let $\bar{a} = (a_1, \ldots, a_n)$. We will be concerned with $\mathrm{tp}(a, b, c/\bar{a})$ and $\mathrm{tp}(a', b', c'/\bar{a})$ (but not especially with the type of a, b, c over a', b', c').

Assume $M_n = \{a_1, \ldots, a_n\}$ have been found, with $\mathrm{tp}(abc/M_n) =_{<3} \mathrm{tp}(a'b'c'/M_n)$. If n is odd we work towards (i), if even towards (ii). Say n is odd. Then we need to find d such that $d, a, b, c, a', b', c', a_1, \ldots, a_n$ is Q-independent, and

(i) $\mathrm{tp}(d, b, c, a_1, \ldots, a_n) = \mathrm{tp}(a, b, c, a_1, \ldots, a_n)$.

(iii) $\mathrm{tp}(a, b, c, /d, a_1 \ldots a_n) =_{<3} \mathrm{tp}(a', b', c'/d, a_1 \ldots a_n)$.

To meet (i), we extend $\mathrm{tp}(a/b, c, \bar{a})$ to a type $p(x, a, b, c, \bar{a})$ over $a, b, c, \bar{a}$, so that $p \in Q$; this will be $\mathrm{tp}(d/a, b, c, \bar{a})$. Next (moving the elements a', b', c' if needed, recalling we are concerned only with their type over $\bar{a}$) we determine a type $\mathrm{tp}(d, a', b', c', \bar{a})$, so that $\mathrm{tp}(a', b', c'/\bar{a}, d) =_{<3} \mathrm{tp}(a, b, c/\bar{a}, d)$; then (iii) is satisfied. This is possible using the induction hypothesis and $(2, \infty)$-amalgamation (or $(2, 3)$-amalgamation over $\bar{a}, d$).

Thus M can be constructed satisfying (i)–(iii). Now the result follows from Lemma B.6. □

Remark B.9. The notion of measure stationarity (the conclusion of Theorem B.8) arose in early work of Elad Levi on the definable higher Szemerédi lemma. Levi observed that it would suffice for a definable version of Gowers' proof of higher-dimensional Szemerédi. For the case of pseudofinite fields, stationarity was eventually proved in the stronger quantitative form, see [29].

But for general theories this remains interesting.

Question B.10. If Y also admits a definable measure commuting with μ, and Fubini is assumed, does stationarity (Theorem B.8) imply a precise formula similar to Theorem B.11, and valid on a set of full measure? It seems plausible that this can be proved by double counting and using Cauchy–Schwarz, as in Theorem 3.16(2).

Also, the proof should extend assuming higher amalgamation holds only for systems of algebraically closed substructures.

Theorem B.11. *Let μ be a definable measure, with Fubini. Let M be a model. Then the measure spaces $S_{x_1,\dots,x_n}(M)$ form an independent system.*

Equivalently, the associated measure algebras $L_{x_1,\dots,x_n}(M)$, with the standard embeddings among them, form an independent system in the usual sense of stability. (See Problem 3.23.)

By a Löwenheim–Skolem argument, we may (and will) assume the language as well as M are countable.

For readability we will write omit the variable letter x, writing $\phi(123)$ for $\phi(x_1, x_2, x_3)$, $\phi(124)$ for $\phi'(x_1, x_2, x_4)$, $L(123)$ for the measure algebra of formulas in x_1, x_2, x_3 over M, $L(12, 23, 13)$ for the join of the measure algebras $L(ij)$ $(1 \le i, j \le 3)$, $S(12, 23, 13)$ for the corresponding (measured) Stone spaces.

Further we let $E(\phi; 12, 13, 23)$ denote the conditional expectation of ϕ relative to $L(12, 23, 13)$.

Over larger structures $M(b)$, $M(c)$ and $M(bc)$, we have the measured Boolean algebras $L(1b)$ of formulas in x_1 over $M(b)$, and likewise $L(1c)$, $L(1bc)$, and $L(1b, 1c)$ generated by $L(1b) \cup L(1c)$; and the Stone spaces $S(1b) = S_{x_1}(M(b))$ and similarly $S(1c)$ and $S(1bc)$.

We view formulas ϕ as $\{0, 1\}$-valued (so conjunction is the same as multiplication), or more generally valued in a bounded interval of $\mathbb{R}$ (so multiplication is still defined).

We use an integral symbol to denote expectation when it is absolute and not conditional; the integral can always be understood to be over the largest space around, such $S(1234)$ when the variables are among x_1, x_2, x_3, x_4. Sometimes we will nevertheless indicate the intended space by a subscript, e.g., $\int_1$ for the integral over $S(1)$.

Proof. The case $n = 4$ is representative. We want then to prove independence of $L(123)$ from $L(124, 134, 234)$ over $L(12, 13, 14)$.

It suffices to prove that

$$\int \phi(123)\phi'(124)\phi''(134)\phi'''(234) = \int E(\phi(123); 12, 13, 23)\phi'(124)\phi''(134)\phi'''(234).$$

This will then extend to all bounded L^1-functions on $S_{123}(M)$ in place of $\phi(123)$; having replaced $\phi(123)$ by $E(\phi(123); 12, 13, 23)$, we can continue and do the same with ϕ', etc. Let $\widehat{\phi} = E(\phi(123); 12, 13, 23)$.

It suffices to prove that for any random triple b, c, d (i.e., $\mathrm{tp}(bcd/M)$ is random), we have $\int_1 \phi(1bc)\phi'(1bd)\phi''(1cd)\phi'''(bcd) = \int_1 \widehat{\phi}(1bc)\phi'(1bd)\phi''(1cd)\phi'''(bcd)$; or, taking out the constant factor $\phi'''(bcd)$, that

$$\int_1 \phi(1bc)\phi'(1bd)\phi''(1cd) = \int_1 \widehat{\phi}(1bc)\phi'(1bd)\phi''(1cd).$$

By Lemma B.1 and Corollary B.2, $\mathrm{tp}(a/Mbcd)$ is finitely satisfiable in M. Thus Lemma B.6 applies, and shows that

$$\int_1 \phi(1bc)\phi'(1bd)\phi''(1cd) = \int_1 E(\phi(1bc); L(1b)L(1c))\phi'(1bd)\phi''(1cd).$$

Thus the equality in the following claim finishes the proof.

Claim. *Let* $\widehat{\phi}(123) = E(\phi(123); 12, 13, 23)$. *Then for random* $\mathrm{tp}(bc/M)$ *we have*

$$\widehat{\phi}(1bc) = E(\phi(1bc); L(1b)L(1c)).$$

Proof. Note that $S(1bc)$ can be identified with the fiber above $\mathrm{tp}(bc/M)$ of $S(123) \to S(23)$, and likewise for $S(1b, 1c, bc)$. Thus the claim follows from Lemma B.4, applied to the maps $S(123) \to S(12, 13, 23) \to S(12)$ and the fibers above $\mathrm{tp}(bc)$. $\square$

Note that the type partition is canonical, once the model is chosen, and approximated by partitions into definable sets.

Since the theorem is only valid over a model, it loses sight of possible symmetries. But if a definable group G acts and is μ-measure preserving, $G(M)$ acts on the type spaces and we do have equivariance.

One can deduce a version of the Hoover–Kallenberg higher-dimensional de Finetti theorem in a similar (or actually easier) way:

Proposition B.12 (Hoover–Kallenberg). *Let μ be a definable measure. Let $\mathbb{N}_1 + \mathbb{N}_2$ be the disjoint sum of two copies of $\mathbb{N}$ ordered by $\mathbb{N}_1 < \mathbb{N}_2$, and let $(a_i : i \in \mathbb{N}_1 + \mathbb{N}_2)$ be an indiscernible sequence. Let $M := \{a_i : i \in \mathbb{N}_1\}$. For $u \subset \mathbb{N} = \mathbb{N}_2$ let S_u be the space of types in variable x over $M \cup \{a_i : i \in u\}$, with measure induced by μ. Then the S_u form an independent system of measure spaces.*

Proof. For $i \in \mathbb{N}$, $\mathrm{tp}(a_i / M \cup \{a_j : j > i\})$ is finitely satisfiable in M. Hence Lemma B.6 applies. □

(Is Fubini needed?)

Remark B.13 (NIP). Assume NIP, and work over a model. Then a statement much stronger than Theorem B.11 holds: Let $\mu(x)$ be a definable measure (with no Fubini self-commutation assumptions). Let B_n be the Boolean algebra of formulas in variables $x_1, \ldots, x_n$, and let $B_{\binom{1}{n}}$ be the subalgebra generated by formulas $\psi(x_i)$ in a single x_i variable. Then for every formula $\phi \in B_n$ and $\epsilon > 0$ there exists $\phi' \in B_{\binom{1}{n}}$ with $\mu(\phi \triangle \phi') < \epsilon$. Equivalently, the induced inclusion of σ-additive measure algebras, up to the null ideals, is an isomorphism. For $n = 2$ this is proved in [48, 1.7(1)], and under slightly different assumptions (essentially Fubini) as Theorem 4.1(a,b) of [60]. (It is curious that while the two teams of authors were entirely unaware of the parallel work in another field, the arXiv submissions are two days apart.) The case of arbitrary n follows immediately by induction from the case $n = 2$. Once one knows that the measure algebra $M(X \times Y)$ is generated by the $M(X)$ and $M(Y)$, it follows that the measure algebra $M(X \times Y \times Z)$ is generated by $M(X \times Y) \cup M(Z)$ and hence by $M(X) \cup M(Y) \cup M(Z)$. With Fubini assumed, a strengthening of this, both quantitative and qualitative, especially for distal theories, is obtained in [26].

All of these sources allow arbitrary parameters. Using definability of the measures, they thus hold over a model.

Question B.14. Does the above strong stationarity for NIP theories, identifying B_n with $B_{\binom{1}{n}}$, hold over bdd(0)? Possibly a statement of this type may follow by the method of Theorem B.8, noting that only the $n = 2$ case is needed and that (2, 3)-amalgamation is obtained in Theorem 3.16.

For an extraordinary generalization to higher-arity NIP, see [27, Corollary 6.10 or Corollary 11.4].

Appendix C. An example from mixing

This appendix to Section 3 is intended to illustrate the use of expectation quantifiers and the various version of the independence Theorem 3.16.

We look at the convolution of two real-valued functions f, g on a group. This is well-studied in connection with mixing (see [58; 32]); I learned about this from a minicourse by Itay Glazer and Emmanuel Breuillard in Oxford in spring 2024.

We will give a simple stability-theoretic proof of a special case of [58, Theorem 1], namely for groups $G(\mathbb{F}_q)$ (modulo center) where G is a simply connected algebraic group. (Using ACFA in place of PF we could also cover Rees and Suzuki group, i.e., the bounded rank case of [58].) Corollary C.8 covers, e.g., nilpotent groups, and was written in response to a question of Glazer's for vector groups; he independently proved the vector group case by analytic means.

Here the simplest version of stationarity (Theorem 2.10) will suffice. But Theorem B.11 would not do; it is valid for almost all pairs of types, but in the proof of Proposition C.1 it is essential to use the same type twice.

Let G be a group carrying a left-invariant definable measure μ. G may include additional relations (of discrete or of continuous logic). We will use pure probability logic quantifiers. Formally, such quantifiers do not directly distinguish the graph of multiplication from $\varnothing$; rather we first define, at the quantifier-free level, binary relations such as $g(t,x) = g(t^{-1}x)$, and only then apply probability quantifiers to such relations.

G may for instance be a compact group made discrete — taken with the discrete metric as a CL structure, and with the Haar measure serving to interpret expectation quantifiers. Assuming the basic relations are measurable, it follows that all formulas obtained by continuous connectives and expectation quantifiers are also measurable. Or G may be an amenable group with a finitely additive invariant measure. But the main example to have in mind will be an ultraproduct of finite groups with their normalized counting measure μ. In this case μ is both left and right invariant. (This will be used in Corollary C.8 to see that the convolution is well-defined and continuous on L^1.)

We begin with the observation that the convolution $f * g$ is definable:

$$f * g(x) = E_t f(t) g(t^{-1}x).$$

We say two real-valued functions h, h' are equal a.e. if $E_x(|h(x) - h'(x)|) = 0$. Let w be an element of G; define $h^w(x) = h(wx)$.

For h a definable function $G \to \mathbb{R}$ define the stabilizer of h to be $\operatorname{Stab}(h) := \{w \in G : h =_{\text{a.e.}} h^w\}$. Then $\operatorname{Stab}(h)$ is an $\bigwedge$-definable subgroup of G.

Proposition C.1. *Let f, g be definable functions into $\mathbb{R}$, and let $h = f * g$ be their convolution. Then the stabilizer of h is a $\bigwedge$-definable subgroup of G of bounded index. In particular if G admits no nontrivial definable homomorphisms into compact groups, then h is constant a.e.*

Proof. Let w be an element of G, and $h = f * g$ the convolution. By left invariance, substituting $w^{-1}t$ for t, we have

$$h^w(x) = E_t(f(t) g(t^{-1}wx)) = E_t(f(w^{-1}t) g(t^{-1}x)).$$

By Proposition 3.8, this is a (real-valued) stable relation between w and x.

Let M_0 be a countable model (taking the bounded closure of 0 will also work). By Theorem 2.10, provided $\mathrm{tp}(a/M_0, b)$ does not divide over M_0 via a stable formula, the value of $h^b(a)$ depends only on $\mathrm{tp}(a/M_0)$ and $\mathrm{tp}(b/M_0)$.

Suppose $\mathrm{tp}(b/M_0) = \mathrm{tp}(c/M_0)$, yet $E_x|h^b(x) - h^c(x)| \geq \epsilon > 0$. By the remarks following Definition 3.10, there exists a such that $|h^b(a) - h^c(a)| \geq \epsilon$ and $\mathrm{tp}(a/M_0(b, c))$ does not divide; in particular $a \perp b, c$ holds, a contradiction. Hence $E_x|h^b(x) - h^c(x)| = 0$, so

$$h^b(x) =_{\text{a.e.}} h^c(x).$$

It follows that $b^{-1}c \in \mathrm{Stab}(h)$. We have shown that any two elements with the same type over M_0 lie in the same coset of $\mathrm{Stab}(h)$; hence $\mathrm{Stab}(h)$ has bounded index. □

Let G_n be a family of groups, endowed with left and right translation invariant finitely additive measures. We say G_n is a quasirandom family, in the sense of Gowers, if for each $d \in \mathbb{N}$, for all sufficiently large n, G_n has no d-dimensional representations. It follows from [30, Theorem 1.1] that if G is any nonprincipal ultraproduct of the G_n, then G admits no nontrivial definable homomorphism into a compact Lie group, and hence by Peter–Weyl no nontrivial definable homomorphism into any compact group. We use this below:

Corollary C.2. *Let (G_n) be a quasirandom family. Let $b > 0$ and let*

$$f, g : G_n \to [-b, b] \subset \mathbb{R}$$

be functions with $\|f\|_1 = \|g\|_1 = 1$. Then $\|f * g - 1\|_1 \to 0$ as $n \to \infty$.

Proof. Set $h = f * g$. By Proposition C.1, in any ultraproduct G, the stabilizer of h is all of G; so h is constant a.e., and since $\|h\|_1 = 1$ the constant value is 1. □

Remark C.3. In the case of bounded rank families of finite simple groups, a simpler proof of Corollary C.2 can be given. Let G be a definably simple group in the theory of pseudofinite fields. It follows from [40, 7.8] that any elementary extension of G is also simple; hence G cannot have a $\bigwedge$-definable subgroup of bounded index *in any expansion to a bigger language.* Thus by Proposition C.1, $f * g$ is constant a.e.

Definition C.4. A connected algebraic group G over $\mathbb{Q}^a$ is called *simply connected* if there is no surjective homomorphism $\widetilde{G} \to G$ of algebraic groups over $\mathbb{Q}^a$ with nontrivial finite kernel.

This definition is usually found in the setting of semisimple groups, but we extend it to all connected algebraic groups. Commutative algebraic groups whose geometric points form a divisible group are clearly simply connected; in particular in characteristic zero. vector groups such as $\mathbb{G}_a^n$ are simply connected.

Corollary C.5. *Let G be a simply connected algebraic group. Let $f, g : G(\mathbb{F}_p) \to \mathbb{R}$ be uniformly definable* (*and hence uniformly bounded*), $\|f\|_1 = \|g\|_1 = 1$. *Then $\|f * g - 1\|_1 \to 0$ as $p \to \infty$.* (*And similarly for prime powers, and for $G(\mathbb{F}_p)/H_p$, where H_p is some uniformly definable normal subgroup of $G(\mathbb{F}_p)$.*)

Proof. Let $h = f * g$. It suffices to show that $F \models E_x(|h - 1|) = 0$ for any ultraproduct F of the finite fields $\mathbb{F}_p$. We have $\|h\|_1 = 1$ so it suffices to show that h is $G(F)$-invariant a.e. This in turn follows from Proposition C.1, once we show G has no proper $\bigwedge$-definable subgroups of bounded index.

Now Theorem 8.5 of [40] shows that $G(F)$ has no definable subgroups of finite index. Theorem 6.3 there says that any $\bigwedge$-definable group H is an intersection of definable groups H_i. If H has bounded index, then each H_i has bounded index and hence by compactness, finite index; but then $H_i = G$ for each i so $H = G$. □

Remark C.6. Unlike Corollary C.2, where arbitrary bounded functions f_n, g_n are allowed, in Corollary C.5 it is essential that they be uniformly definable over finite fields.

For instance on $\mathbb{G}_a$, if $f_p(x \bmod p) = 10$ for $0 \leq x < p/10$, and $f_p(u) = 0$ for all other $u \in \mathbb{F}_p$, we see that the ultraproduct f requires at least 10 self-convolutions to become uniform, and not 2. This accounts for the additional model-theoretic ingredient (Theorems 8.5 and 6.3) quoted above; the homomorphism $n \bmod p \mapsto \exp(2\pi i/p)$ exists, and must be shown not to be uniformly definable.

To connect to convolution of pushforward measures, we will need one simple geometric lemma:

Lemma C.7. *Let $f : Y \to X$ be a dominant morphism of irreducible varieties over an ultrapower $F = \lim_u \mathbb{F}_q$ of finite fields $\mathbb{F}_q$. Let d be the generic fiber dimension, and $d_X = \dim(X)$. Let $F(x) = q^{-d}|f^{-1}(x)|$. Then there exists a definable* (*in* $\mathrm{Th}(F)$) *real-valued function f on X and a proper subvariety X_0 of X, such that $f - F$ tends to 0 uniformly along u on $X \setminus X_0$. Also $\|F - f\|_1$ tends to 0.*

Proof. Up to the last sentence, this is the main result of [23]. We may take f to vanish on X_0. The last sentence follows since $Y_0 := f^{-1}(X_0)$ is a proper subvariety of Y and hence has dimension $< \dim(Y)$. So

$$q^{-\dim(X)} \sum_{x \in X_0} F(x) = q^{\dim(Y)}|Y_0| = O(q^{-1/2}). \qquad \square$$

Corollary C.8. *Let G be a simply connected algebraic group defined over $\mathbb{Z}[m^{-1}]$, and let $\eta_i : Y_i \to G$ be a dominant morphism of irreducible varieties* ($i = 1, 2$). *Let ν_p^i be the normalized counting measure on $Y_i(\mathbb{F}_p)$ and let $\mu_p^i = (\eta_i)_* \nu_p^i$ be the pushforward of ν_p^i to G. Also for a prime $p > m$ let μ_p be Haar on $G_p(\mathbb{F}_p)$. Then $\|\mu_p - \mu_p^1 * \mu_p^2\|_1 \to 0$ as $p \to \infty$.*

Proof. We may write $\mu_p^i = F_i \mu_p$. Let f_i be as in Lemma C.7, so $\|\mu_p^i - f_i \mu_p\|_1 \to 0$. Now the statement follows using continuity of convolution on the L^1-norm. □

To compare this to [58, Theorem 1], set $Y_i = G^{d_i}$, where $w_i = w_i(x_1, \dots, x_{d_i})$, and let f_i be the word map; then $f_1 * f_2$ is the word map associated to $w_1 w_2$, since they have disjoint variables.

Acknowledgements

This text is largely based on notes for a course given in Paris in Fall 2015, under the auspices of the Fondation Sciences Mathématique de Paris, and retains an expository aspect in the first sections. I am very grateful to the participants and to the FSMP for their support. Thanks also to Emmanuel Breuillard and Alex Gamburd for many interesting conversations on the subjects of this paper, in particular at the time of the Groups and Words meeting when Gamburd was working on [15]. Section 6 was motivated by these, as a kind of soft background result. Thanks to Alexis Chevalier and Mira Tartarotti for their comments on Section 3, Pierre Simon on a previous version of Appendix A, and to Shahar Mozes and Benjy Weiss for conversations around Keisler's relational law of large numbers. And finally thanks to Dan Drai for many conversations that brought out, in particular, the relation of the stabilizer theorem or the graphon metric to measures used in practice in natural language processing.

The referees of this paper deserve exceptional credit. In addition to many local comments and references that greatly contributed to the paper, a serious issue was uncovered in both Section 5 and Section 6 (and in a slightly different form in Section 7). In both cases, the structure of approximate equivalence relation is precisely described *piecewise*, but only a partial account is given of the interaction between the pieces. See in particular Problem 5.11 and Section 6.12.

References

[1] N. Avni and S. Garion, "Connectivity of the product replacement graph of simple groups of bounded Lie rank", *J. Algebra* **320**:2 (2008), 945–960. MR Zbl

[2] B. Bakker, B. Klingler, and J. Tsimerman, "Tame topology of arithmetic quotients and algebraicity of Hodge loci", *J. Amer. Math. Soc.* **33**:4 (2020), 917–939. MR Zbl

[3] I. Ben-Yaacov, "Schrödinger's cat", *Israel J. Math.* **153** (2006), 157–191. MR Zbl

[4] I. Ben Yaacov, "Continuous first order logic for unbounded metric structures", *J. Math. Log.* **8**:2 (2008), 197–223. MR Zbl

[5] I. Ben Yaacov, "Continuous and random Vapnik–Chervonenkis classes", *Israel J. Math.* **173** (2009), 309–333. MR Zbl

[6] I. Ben Yaacov, "Model theoretic stability and definability of types, after A. Grothendieck", *Bull. Symb. Log.* **20**:4 (2014), 491–496. MR Zbl

[7] I. Ben-Yaacov and A. Berenstein, "Imaginaries in Hilbert spaces", *Arch. Math. Logic* **43**:4 (2004), 459–466. MR Zbl

[8] I. Ben Yaacov and A. Usvyatsov, "Continuous first order logic and local stability", *Trans. Amer. Math. Soc.* **362**:10 (2010), 5213–5259. MR Zbl

[9] I. Ben Yaacov, A. Berenstein, C. W. Henson, and A. Usvyatsov, "Model theory for metric structures", pp. 315–427 in *Model theory with applications to algebra and analysis*, vol. 2, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **350**, Cambridge Univ. Press, 2008. MR Zbl

[10] I. Benjamini and T. Hutchcroft, "Large, lengthy graphs look locally like lines", *Bull. Lond. Math. Soc.* **53**:2 (2021), 482–492. MR Zbl

[11] I. Benjamini and O. Schramm, "Recurrence of distributional limits of finite planar graphs", *Electron. J. Probab.* **6** (2001), art. id. 23. MR Zbl

[12] I. Benjamini, H. Finucane, and R. Tessera, "On the scaling limit of finite vertex transitive graphs with large diameter", *Combinatorica* **37**:3 (2017), 333–374. MR Zbl

[13] N. Bourbaki, *Éléments de mathématique*: *Variétés différentielles et analytiques*, *fascicule de résultats*, Springer, 1981.

[14] J. Bourgain, A. Gamburd, and P. Sarnak, "Affine linear sieve, expanders, and sum-product", *Invent. Math.* **179**:3 (2010), 559–644. MR Zbl

[15] J. Bourgain, A. Gamburd, and P. Sarnak, "Markoff triples and strong approximation", *C. R. Math. Acad. Sci. Paris* **354**:2 (2016), 131–135. MR Zbl

[16] S. Braunfeld, C. Jahel, and P. Marimon, "When invariance implies exchangeability (and applications to invariant Keisler measures)", forthcoming.

[17] E. Breuillard, "Lectures on approximate groups and Hilbert's 5th problem", pp. 369–404 in *Recent trends in combinatorics*, IMA Vol. Math. Appl. **159**, Springer, 2016. MR Zbl

[18] E. Breuillard, B. Green, and T. Tao, "The structure of approximate groups", *Publ. Math. Inst. Hautes Études Sci.* **116** (2012), 115–221. MR Zbl

[19] S. Cantat and F. Loray, "Dynamics on character varieties and Malgrange irreducibility of Painlevé VI equation", *Ann. Inst. Fourier* (*Grenoble*) **59**:7 (2009), 2927–2978. MR Zbl

[20] E. Casanovas, D. Lascar, A. Pillay, and M. Ziegler, "Galois groups of first order theories", *J. Math. Log.* **1**:2 (2001), 305–319. MR Zbl

[21] C. C. Chang and H. J. Keisler, *Continuous model theory*, Annals of Mathematics Studies **58**, Princeton University Press, 1966. MR Zbl

[22] Z. Chatzidakis and E. Hrushovski, "Model theory of difference fields", *Trans. Amer. Math. Soc.* **351**:8 (1999), 2997–3071. MR Zbl

[23] Z. Chatzidakis, L. van den Dries, and A. Macintyre, "Definable sets over finite fields", *J. Reine Angew. Math.* **427** (1992), 107–135. MR Zbl

[24] G. Cherlin and E. Hrushovski, *Finite structures with few types*, Annals of Mathematics Studies **152**, Princeton University Press, 2003. MR Zbl

[25] A. Chernikov and P. Simon, "Definably amenable NIP groups", *J. Amer. Math. Soc.* **31**:3 (2018), 609–641. MR Zbl

[26] A. Chernikov and S. Starchenko, "Definable regularity lemmas for NIP hypergraphs", *Q. J. Math.* **72**:4 (2021), 1401–1433. MR Zbl

[27] A. Chernikov and H. Towsner, "Hypergraph regularity and higher arity VC-dimension", preprint, 2020. arXiv 2010.00726

[28] A. Chevalier and E. Hrushovski, "Piecewise interpretable Hilbert spaces", preprint, 2021. arXiv 2110.05142

[29] A. Chevalier and E. Levi, "An algebraic hypergraph regularity lemma", preprint, 2022. arXiv 2204.01158

[30] G. Conant, E. Hrushovski, and A. Pillay, "Compactifications of pseudofinite and pseudo-amenable groups", preprint, 2023. arXiv 2308.08440

[31] L. van den Dries, "Dimension of definable sets, algebraic boundedness and Henselian fields", *Ann. Pure Appl. Logic* **45**:2 (1989), 189–209. MR Zbl

[32] I. Glazer and Y. I. Hendel, "On singularity properties of convolutions of algebraic morphisms — the general case", *J. Lond. Math. Soc.* (2) **103**:4 (2021), 1453–1479. MR Zbl

[33] I. Goldbring and H. Towsner, "An approximate logic for measures", *Israel J. Math.* **199**:2 (2014), 867–913. MR Zbl

[34] W. T. Gowers and J. Long, "Partial associativity and rough approximate groups", *Geom. Funct. Anal.* **30**:6 (2020), 1583–1647. MR Zbl

[35] M. Gromov, "Structures, learning and ergosystems", draft book, 2011, available at https://www.ihes.fr/~gromov/category/ergosystems.

[36] A. Grothendieck, "Critères de compacité dans les espaces fonctionnels généraux", *Amer. J. Math.* **74** (1952), 168–186. MR Zbl

[37] B. Hart, B. Kim, and A. Pillay, "Coordinatisation and canonical bases in simple theories", *J. Symbolic Logic* **65**:1 (2000), 293–309. MR Zbl

[38] D. Haussler and E. Welzl, "ϵ-nets and simplex range queries", *Discrete Comput. Geom.* **2**:2 (1987), 127–151. MR Zbl

[39] E. Hrushovski, "Simplicity and the Lascar group", preprint, 1998, available at http://people.maths.ox.ac.uk/hrushovski/preprints.html.

[40] E. Hrushovski, "Pseudo-finite fields and related structures", pp. 151–212 in *Model theory and applications*, Quad. Mat. **11**, Aracne, Rome, 2002. MR Zbl

[41] E. Hrushovski, "Beyond the Lascar group", preprint, 2011. arXiv 2011.12009

[42] E. Hrushovski, "Groupoids, imaginaries and internal covers", *Turkish J. Math.* **36**:2 (2012), 173–198. MR Zbl

[43] E. Hrushovski, "Stable group theory and approximate subgroups", *J. Amer. Math. Soc.* **25**:1 (2012), 189–243. MR Zbl

[44] E. Hrushovski and A. Pillay, "Definable subgroups of algebraic groups over finite fields", *J. Reine Angew. Math.* **462** (1995), 69–91. MR Zbl

[45] E. Hrushovski and A. Pillay, "On NIP and invariant measures", *J. Eur. Math. Soc.* **13**:4 (2011), 1005–1061. MR Zbl

[46] E. Hrushovski and F. Wagner, "Counting and dimensions", pp. 161–176 in *Model theory with applications to algebra and analysis*, vol. 2, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **350**, Cambridge Univ. Press, 2008. MR Zbl

[47] E. Hrushovski, Y. Peterzil, and A. Pillay, "Groups, measures, and the NIP", *J. Amer. Math. Soc.* **21**:2 (2008), 563–596. MR Zbl

[48] E. Hrushovski, A. Pillay, and P. Simon, "Generically stable and smooth measures in NIP theories", *Trans. Amer. Math. Soc.* **365**:5 (2013), 2341–2366. MR Zbl

[49] T. Ibarlucía, "Infinite-dimensional Polish groups and Property (T)", *Invent. Math.* **223**:2 (2021), 725–757. MR Zbl

[50] C. Jahel, *Des progrès sur le problème d'unique ergodicité*, Ph.D. thesis, Université Claude Bernard Lyon 1, 2023, available at https://theses.hal.science/tel-03856013.

[51] C. Jahel and T. Tsankov, "Invariant measures on products and on the space of linear orders", *J. Éc. Polytech. Math.* **9** (2022), 155–176. MR Zbl

[52] H. J. Keisler, "Hyperfinite model theory", pp. 5–110 in *Logic Colloquium* 76 (Oxford, 1976), Stud. Logic Found. Math. **87**, North-Holland, Amsterdam, 1977. MR Zbl

[53] H. J. Keisler, "Probability quantifiers", pp. 509–556 in *Model-theoretic logics*, edited by J. Barwise and S. Feferman, Springer, 1985. MR Zbl

[54] B. Kim and A. Pillay, "Simple theories", *Ann. Pure Appl. Logic* **88**:2-3 (1997), 149–164. MR Zbl

[55] B. Kim and A. Pillay, "From stability to simplicity", *Bull. Symbolic Logic* **4**:1 (1998), 17–36. MR Zbl

[56] S. Kobayashi, *Transformation groups in differential geometry*, Ergebnisse der Math. **70**, Springer, 1972. MR Zbl

[57] J.-L. Krivine and B. Maurey, "Espaces de Banach stables", *Israel J. Math.* **39**:4 (1981), 273–295. MR Zbl

[58] M. Larsen, A. Shalev, and P. H. Tiep, "Probabilistic Waring problems for finite simple groups", *Ann. of Math.* (2) **190**:2 (2019), 561–608. MR Zbl

[59] L. Lovász, *Large networks and graph limits*, Amer. Math. Soc. Colloquium Publ. **60**, American Mathematical Society, Providence, RI, 2012. MR Zbl

[60] L. Lovász and B. Szegedy, "Regularity partitions and the topology of graphons", pp. 415–446 in *An irregular mind*: *Szemerédi is* 70, Bolyai Soc. Math. Stud. **21**, János Bolyai Math. Soc., Budapest, 2010. MR Zbl

[61] L. Lovász and B. Szegedy, "The automorphism group of a graphon", *J. Algebra* **421** (2015), 136–166. MR Zbl

[62] D. Macpherson and K. Tent, "Profinite groups with NIP theory and p-adic analytic groups", *Bull. Lond. Math. Soc.* **48**:6 (2016), 1037–1049. MR Zbl

[63] P. Marimon, "Invariant Keisler measures for ω-categorical structures", preprint, 2022. arXiv 2211.14628

[64] P. Marimon, "On the non-measurability of ω-categorical Hrushovski constructions", preprint, 2022. arXiv 2208.06323

[65] P. Marimon, *Measures and amalgamation properties in ω-categorical structures*, Ph.D. thesis, Imperial College London, 2023, available at https://spiral.imperial.ac.uk/handle/10044/1/106470.

[66] C. R. Matthews, L. N. Vaserstein, and B. Weisfeiler, "Congruence properties of Zariski-dense subgroups, I", *Proc. London Math. Soc.* (3) **48**:3 (1984), 514–532. MR Zbl

[67] M. V. Nori, "On subgroups of $\mathrm{GL}_n(\mathbb{F}_p)$", *Invent. Math.* **88**:2 (1987), 257–275. MR Zbl

[68] A. Pillay, "Type-definability, compact Lie groups, and o-minimality", *J. Math. Log.* **4**:2 (2004), 147–162. MR Zbl

[69] A. A. Razborov, "Flag algebras", *J. Symbolic Logic* **72**:4 (2007), 1239–1282. MR Zbl

[70] T. Sanders, "On a nonabelian Balog–Szemerédi-type lemma", *J. Aust. Math. Soc.* **89**:1 (2010), 127–132. MR Zbl

[71] S. Shelah, "The lazy model-theoretician's guide to stability", *Logique et Anal.* (*N.S.*) **18**:71-72 (1975), 241–308. MR Zbl

[72] T. Tao, *Hilbert's fifth problem and related topics*, Graduate Studies in Mathematics **153**, American Mathematical Society, Providence, RI, 2014. MR Zbl

[73] H. Towsner, "A model theoretic proof of Szemerédi's theorem", preprint, 2010. arXiv 1002.4456

[74] V. N. Vapnik and A. J. Chervonenkis, "The uniform convergence of frequencies of the appearance of events to their probabilities", *Theor. Prob. Appl.* **16**:2 (1971), 264–280. MR Zbl

[75] A. M. Vershik, "Random and universal metric spaces", pp. 199–228 in *Dynamics and randomness*, vol. II, Nonlinear Phenom. Complex Systems **10**, Kluwer, Dordrecht, 2004. MR Zbl

[76] H. Yamabe, "On the conjecture of Iwasawa and Gleason", *Ann. of Math.* (2) **58** (1953), 48–54. MR Zbl

Received 15 Dec 2022. Revised 23 Jun 2024.

EHUD HRUSHOVSKI:
ehud.hrushovski@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.417

Independence and bases: theme and variations

Peter J. Cameron

To Boris Zilber on the occasion of his 75th birthday.

This paper describes a complex of related ideas, ranging from Urbanik's v^*-*algebras*, through Deza's *geometric groups* and Zilber's *homogeneous geometries*, to Sims' *bases* for permutation groups and their use in defining "size" parameters on finite groups, with a brief look at Cherlin's *relational complexity*. It is not a complete survey of any of these topics, but aims to describe the links between them.

1. Introduction

In the 1980s, there was widespread interest in matroids with a large amount of symmetry. Michel Deza was studying *perfect matroid designs*, matroids in which the cardinality of a flat depends only on its dimension: this class includes uniform matroids, classical projective and affine spaces, and Steiner systems. One way to enforce this condition is to assume a large group of automorphisms: for example, it holds if the stabiliser of any subset is transitive on the set of points not dependent on that subset. The tool of choice for many was the recently announced classification of finite simple groups.

In 1988, I attended a Durham Symposium on model theory and groups run by the London Mathematical Society. To my amazement, Boris Zilber spoke at the symposium, giving four lectures on his recent result classifying such geometries with rank at least seven, by geometric methods not using CFSG.

Zilber [1984] had worked on first-order theories which are categorical in all cardinalities. We knew from Morley's theorem that this imposes just two conditions on the theory: countable and uncountable categoricity. The result of Engeler, Ryll-Nardzewski and Svenonius shows that $\aleph_0$-categoricity is equivalent to the existence of a large automorphism group, while categoricity in higher powers forces structural conditions such as the existence of rank functions, which led to the development of stability theory. These nicely combine if both types of categoricity hold. In

MSC2020: primary 20B10; secondary 03C35, 05B35.
Keywords: strictly minimal structures, bases, independence algebras.

particular, strictly minimal countably categorical theories carry geometries of infinite dimension which have analogues of the properties that Deza was interested in.

Zilber's achievement in his lectures at the symposium (described in [Zilber 1988a]) was to observe that methods from the infinite case could be applied also to finite structures. The set of elements independent of a finite subset is infinite in the infinite case, but sufficiently large in the finite case that arguments can be adapted.

Perhaps Zilber's methods have not been sufficiently integrated into finite combinatorics; we still have work to do.

2. Definitions

A family $\mathcal{B}$ of finite subsets of a set is said to have the *exchange property* if, given $B_1, B_2 \in \mathcal{B}$ and $y \in B_2 \setminus B_1$, there exists $x \in B_1 \setminus B_2$ such that $(B_1 \setminus \{x\}) \cup \{y\} \in \mathcal{B}$.

Clearly all the sets in such a family have the same cardinality. One definition of a *matroid*, in terms of its set of bases, is as a collection of subsets of a finite set having the exchange property. Matroids form an important class of structures, describing subsets of vector spaces, edge sets of graphs (where the bases are spanning forests), transversals to families of sets, and several others. Areas of mathematics in which matroids occur include algebraic and tropical geometry and homotopy theory [Giansiracusa and Giansiracusa 2018].

The set of bases in a vector space V has two properties, which serve as a foundation for linear algebra:

(a) it has the exchange property;

(b) any map from a basis into V has a unique extension to an endomorphism of V.

It is natural to look for further examples of this phenomenon.

Let A be an algebra, in the sense of universal algebra; that is, a set carrying a number of operations of various arities (we interpret 0-ary operations as constants). Suppose that A is finitely generated, and let $\mathcal{B}$ be the set of minimal (under inclusion) generating sets for A. Then A is an *independence algebra* if $\mathcal{B}$ has the above two properties. The definition, and the classification of these algebras, are essentially due to Kazimierz Urbanik [1966] (who called them v^*-*algebras*), but his work was not as well known as it deserved to be, and the concept was later rediscovered in the context of semigroup theory. Section 2 of this paper gives some details about independence algebras and their classification, and mentions a recent result on them.

Perhaps unaware of the work of Urbanik, semigroup theorists Fountain and Lewin [1992] and Gould [1995] realised that earlier structural results of Howie, Reynolds and Sullivan, and Erdos on the full transformation semigroup and the semigroup of linear maps on a vector space could be generalised to endomorphism semigroups of independence algebras. I learned of the topic from John Fountain, and set to work with Csaba Szabó to classify at least the finite independence algebras.

The endomorphisms of a structure form a monoid, and the automorphisms form its group of units. The class of permutation groups arising as automorphism groups of independence algebras is part of a more general class, named "geometric groups" by Michel Deza [Cameron and Deza 1979]. These groups, and the underlying closure systems, were studied by Boris Zilber in the 1980s, in the course of his important researches on countably categorical and $\aleph_0$-stable first-order structures; he called these objects *quasi-Urbanik structures*; see [Zilber 1988b]. Section 3 of this paper discusses some of their theory.

The concept of a *base* can be defined for any permutation group, not just the geometric groups. Permutation group bases do not usually satisfy the exchange property; those which do, the so-called *IBIS groups*, introduced by Dima Fon-Der-Flaass and me [Cameron and Fon-Der-Flaass 1995], form a very interesting class. Bases were introduced by Charles Sims [1970] for use in computation with permutation groups, but raise various interesting questions; among other uses, they form part of László Babai's work on the graph isomorphism problem [Babai 2015], and are connected with Cherlin's notion of *relational complexity* for studying finite homogeneous structures [Cherlin 2016; Cherlin et al. 1996]. This area is seeing renewed activity at present, so I will not give a complete survey, but will highlight some open questions.

The final section draws connections between various "measures" of a finite group with parameters defined in terms of bases in all actions of the group. There are a number of open problems here.

To conclude this section, I give a couple of essential definitions for permutation group theory. Let G be a permutation group acting on Ω. The action is *transitive* if there is no G-invariant subset except for Ω and the empty set (equivalently, any element of Ω can be mapped to any other by some element of G); it is *primitive* if it is transitive and, in addition, there is no G-invariant partition except for the partition into singletons and the partition with a single part. The *stabiliser* of an element $\alpha \in \Omega$ is the subgroup of G consisting of all elements mapping α to itself. The action is *semiregular* or *free* if the stabiliser of every point is the identity; it is *regular* if it is transitive and semiregular.

3. Independence algebras

An *independence algebra* is a finitely generated algebra with the properties

(a) the minimal generating sets have the exchange property;

(b) any map from a minimal generating set into the algebra extends uniquely to an endomorphism of the algebra.

The definition makes no explicit mention of the operations of the algebra; there is some freedom about these, as long as the correct subalgebras and endomorphisms are obtained. (A minimal generating set is a subset minimal with respect to being contained in no proper subalgebra.) Thus a classification up to isomorphism is not possible. Urbanik classified the algebras up to *clone equivalence*, noting that an algebra clone-equivalent to an independence algebra is an independence algebra. Szabó and I used a slightly weaker, but arguably more natural, equivalence. We say that two algebras A and B are *SE-equivalent* if there is a bijection between them which preserves subalgebras and endomorphisms. It turns out that there is one case only where these notions differ (one SE-equivalence class of independence algebras splits into two clone-equivalence classes). For further discussion see [Araújo et al. 2022; 2011; Araújo and Fountain 2004].

There is another small difference also: Urbanik did not allow constants, but used constant-valued unary operations instead. We will see that the presence or absence of constants is crucial to the classification.

First consider the case where A has rank 1. (The rank is the cardinality of a minimal generating set; the exchange property guarantees its invariance.) If there are no constants, then any singleton subset of A is a generating set, and any element can be mapped to any other by a unique automorphism. Thus the automorphism group G of A acts regularly. For any group G, there is an independence algebra of this form; we take $A = G$ and, for each $g \in G$, equip A with a unary operation μ_g given by $\mu_g(x) = gx$. Then G acts regularly on A by right multiplication.

Now suppose that there is a set C of constants. Then the bases are the singletons not contained in C, and any of them can be mapped to any other by a unique automorphism; so we can identify $A \setminus C$ with a group G. There is a unique endomorphism f_c mapping the identity element of G to c, for each $c \in C$. There is a left action of G on C, defined by the rule that the endomorphism $g \circ f_c$ (composed left-to-right) maps the identity of G to an element which we take to be $g(c)$. Conversely, given any group G and left action of G on a set C we obtain an independence algebra on the disjoint union $G \cup C$: it has a constant γ_c for each $c \in C$ (interpreted as c) and a unary operation μ_g for each $g \in G$ given by $\mu_g(h) = gh$ for $h \in G$, $\mu_g(c) = g(c)$.

The phrase "an independence algebra of rank 1" seems to me a remarkably concise way of defining a group with an action on a set.

This construction extends as follows. For any set X, and any group G with a left action on C, define an algebra on the set $A = (X \times G) \cup C$ with C as set of constants, and unary operations μ_g defined by

$$\mu_g((x, h)) = (x, gh), \qquad \mu_g(c) = g(c).$$

This is an independence algebra whose subalgebras are all the sets $(Y \times G) \cup C$

for $Y \subseteq X$, so that the subalgebra lattice is the Boolean algebra $B(X)$ of subsets of X. Every finitely generated independence algebra whose subalgebra lattice is a Boolean algebra is of this form; these are the *trivial* independence algebras. (We could follow the model theorists and call them *disintegrated*.)

Next, it is shown that the subalgebra lattice of a nontrivial independence algebra is a projective or affine space, depending on whether the algebra has constants or not. The arguments for this are quite general, not assuming finiteness or even finite rank; an accessible account is in [Cameron and Szabó 2000].

Finally it is shown that the algebras are of three types:

(a) Let V be a vector space over a division ring F. Then there is an independence algebra whose elements are those of V; the operations are addition in V and scalar multiplication by elements of F. The subalgebras are the subspaces of V, and so the subalgebra lattice is the projective space built on V. If W is a subspace of V, we obtain an independence algebra by taking the elements of W to be constants; its subalgebra lattice is the projective space on V/W.

(b) Let V be a vector space over a division ring F. For each $c \in F$ with $c \neq 0, 1$, define a binary operation $\beta_c(x, y) = cx + (1-c)y$. (If $|F| = 2$, we use instead the ternary operation $\tau(x, y, z) = x + y + z$.) This defines an independence algebra whose subalgebras are the affine subspaces. If W is a subspace of V, we can add unary operations for translations by elements of W to obtain an algebra whose subalgebras are the unions of cosets of W corresponding to affine subspaces of V/W.

(c) Let G be a *sharply* 2*-transitive group* on Ω: this means that any pair of distinct elements of Ω can be mapped to any other such pair by a unique element of G. Let $\{O_i : i \in I\}$ be the set of orbits of G on triples of distinct elements of Ω. For each $i \in I$, define a binary operation μ_i by

$$\mu_i(a, a) = a, \qquad \mu_i(a, b) = c \text{ if } (a, b, c) \in O_i.$$

This defines an independence algebra; its endomorphisms are the elements of G together with the constant functions, and its rank is 2.

The sharply 2-transitive groups have been of interest for a long time, partly because of their connections with projective planes. It was known to Burnside and Frobenius, and probably earlier, that a finite sharply 2-transitive group has a regular normal subgroup, and so is the group of 1-dimensional affine maps over a finite *nearfield* (this is a structure satisfying the field axioms except possibly the commutativity of multiplication and one distributive law). The finite nearfields were all determined by Zassenhaus [1935]: there are infinitely many *Dickson nearfields* (obtained from Galois fields by "twisting" the multiplication) and seven *exceptional nearfields*.

For a long time it was not known whether all sharply 2-transitive groups are given by nearfields. Several authors [Grätzer 1963; Kerby 1974; Tits 1952; Wilke 1972] defined algebraic structures from such groups, which were given various names and satisfied slightly different sets of axioms. Eventually the question was resolved in [Rips et al. 2017]: infinite sharply 2-transitive groups do not necessarily have regular normal subgroups, and so cannot all be defined from nearfields.

The most recent appearance of independence algebras is in the paper [Araújo et al. 2022]. This gives more details about the relation between Urbanik's v^*-algebras and independence algebras, the relation between clone equivalence and SE-equivalence, and the classification theorem, and goes on to develop matrix theory for most types of independence algebras, though in the case of sharply 2-transitive groups this works only for those defined over nearfields.

4. Geometric groups

The automorphism group of an independence algebra has some remarkable properties:

(a) it acts transitively on (ordered) bases for the algebra;

(b) the stabiliser of any tuple of points fixes pointwise the subalgebra they generate and acts transitively on the points outside this subalgebra.

Forgetting the algebra, the problem of determining the groups with properties like these arises in a couple of places:

- The automorphism group of a strictly minimal set in a totally categorical first-order structure has this property, where the role of subalgebras is taken by definably closed sets; as we saw, Boris Zilber was motivated by this.
- Michel Deza defined an analogue of a matroid in the semilattice of partial permutations rather than the lattice of subsets; he called it a permutation geometry (by analogy with the term "combinatorial geometry", an alternative name for a matroid, proposed by Gian-Carlo Rota).

I will not describe the motivation further, but go straight to the definition. These groups were called "geometric groups" by Deza; not an ideal name, since there are many ways in which a group can be "geometric", and there is no connection with the topic of geometric group theory, but I will stick to this term.

A *geometric group*, then, is a permutation group in which the stabiliser of any finite tuple acts transitively on the points it does not fix (if any).

We see immediately that, in a geometric group, the analogue of a basis for independence algebras can be defined as a sequence of points $(x_1, x_2, \dots, x_r)$ in which each point is moved by the stabiliser of its predecessors, but the stabiliser of the whole sequence is the identity. Then the group acts transitively on ordered bases. The number r is the *rank* of the geometric group.

What are the geometric groups? It is clear that a geometric group of rank 1 is an arbitrary group acting regularly, perhaps with some added fixed points. So we can assume that the rank is at least 2.

As noted above, Zilber [1984] determined all geometric groups of rank at least 7: they are stabilisers of sequences of points in the symmetric group, the general linear group, or the affine group (the last two over a finite field). His proof used elementary arguments inspired by model theory. To elaborate a little, [Zilber 1984] analysed the structure of countably categorical $\aleph_0$-stable structures via their strongly minimal sets, showing as a result that totally categorical structures could not be finitely axiomatised. Strictly minimal sets in these structures involve locally finite geometries which are shown to be either disintegrated (all subsets of a set) or projective or affine spaces over finite fields; it is this result which he was able to "finitise", giving the characterisation noted at the start of this paragraph.

At about the same time, Maund [1989] used the recently announced classification of finite simple groups to determine all geometric groups of rank at least 2. The bulk of the work is involved in determining those groups of rank 2, since they occur as building blocks for the groups of larger rank. The list is as follows:

(a) $H \wr S_2$, where H acts regularly.

(b) $M \cdot S_3 \le H \wr S_3$, where H is abelian and regular and

$$M = \{(h_1, h_2, h_3) \in H^3 : h_1 h_2 h_3 = 1\}.$$

(c) Sharply 2-transitive groups.

(d) $V^2 \cdot \mathrm{AGL}(1, q)$ or $V^2 \cdot \mathrm{GL}(2, q)$, where V is a vector space over $\mathrm{GF}(q)$.

(e) $C_{(q-1)/2} \times \mathrm{PSL}(2, q)$ with $q \equiv 3 \pmod 4$.

(f) $C_{q-1} \times \mathrm{Sz}(q)$, where q is an odd power of 2.

(g) $\mathrm{PGL}(3, 2)$ and $\mathrm{PGL}(3, 3)$.

(In case (d), we regard V^2 as $V \otimes W$, where $\dim(W) = 2$, and $\mathrm{GL}(2, q)$ or its subgroup $\mathrm{AGL}(1, q)$ acts on W.)

Maund used this list and some geometry to determine all finite geometric groups of rank at least 2. Unfortunately this work has never been published.

This list was used in [Cameron and Szabó 2000] to give a determination of finite independence algebras. For each geometric group we have to decide whether or not it is possible to define maps to play the role of endomorphisms, and operations preserved by the group to make the domain into an algebra.

5. Bases for permutation groups

The concept of a base for a permutation group arose in computational group theory. A *base* is a sequence of points in the permutation domain whose pointwise stabiliser is the identity. Thus, for geometric groups, bases in the sense previously defined are bases here also.

The importance of a base is that two elements of a permutation group G which agree on a base for G must coincide: for if g and h are the two elements, then gh^{-1} fixes the base pointwise, so $gh^{-1} = 1$. This can lead to a compact representation of group elements if the base size is small. So it is of interest to find a small base for a permutation group. Let $b_m(G)$ be the size of a smallest base for G.

We can find a base very simply, by choosing points and stabilising them until we reach the identity. This is potentially rather wasteful. Though it is hard to find the base of smallest size for a given group, there are two simple methods which perform rather well, involving choosing base points in order:

- There is no need to include a point which is fixed by the stabiliser of the points already chosen. We call a base *irredundant* if no point is fixed by the stabiliser of its predecessors. We note that bases of geometric groups in the earlier sense are by definition irredundant.
- Motivated by this, a good heuristic is to choose each new base point from an orbit of largest size of the stabiliser of its predecessors. This is a "greedy algorithm", and a base produced by this algorithm is called a *greedy* base. The heuristic is based on the idea that to descend a chain of subgroups to the identity, we should choose the subgroup of largest possible index in its predecessor at each stage, and the index of the stabiliser of a point is the size of the orbit of that point.

Note that bases are ordered sequences, and there is no guarantee that reordering an irredundant or greedy base will result in another with the same property.

Clearly, for a geometric group, irredundant bases and greedy bases are the same, and they have a beautiful geometric structure: they are the bases of a matroid. However, the last condition holds more generally, according to this remarkable theorem of Cameron and Fon-Der-Flaass [1995].

Theorem 1. *For a permutation group G, the following conditions are equivalent*:

(a) *all irredundant bases have the same size*;

(b) *the result of reordering an irredundant base is still irredundant*;

(c) *the irredundant bases are the bases of a matroid.*

Proof. Clearly (c) implies (a). Also (a) implies (b), since if reordering a base created a redundancy then a smaller irredundant base could be obtained by removing some elements. Suppose that (b) holds, and let $(a_1, \ldots, a_r)$ and $(b_1, \ldots, b_s)$ be irredundant bases. The stabiliser of $a_1, \ldots, a_{r-1}$ cannot fix all of $b_1, \ldots, b_s$; suppose that it moves b_i. Then $(a_1, \ldots, a_{r-1}, b_i, a_r)$ is a base, which must be redundant since swapping the last two elements gives a redundant base. But the only possible redundancy is that a_r is fixed by the stabiliser of the earlier points, so $(a_1, \ldots, a_{r-1}, b_i)$ is an irredundant base. Thus the exchange property holds. □

Groups satisfying these conclusions are called *IBIS groups* (an acronym for "*I*rredundant *B*ases of *I*nvariant *S*ize". Every geometric group is an IBIS group; the converse is far from true. For a simple example, a *Frobenius group* (a transitive group in which the stabiliser of any two points is trivial but the stabiliser of a single point is not) is an IBIS group of rank 2: the bases are all the 2-element sets. A Frobenius group is a geometric group if and only if it is sharply 2-transitive, and as we saw, all sharply 2-transitive groups are automorphism groups of independence algebras.

A large class of (intransitive) examples is given by the following construction.

Let C be a linear code of length n over the finite field F (a subspace of F^n). Let G be the additive group of C, and let $\Omega = \{1, \ldots, n\} \times F$. Define an action of G on Ω by

$$a : (i, x) \mapsto (i, x + a_i)$$

for $a = (a_1, \ldots, a_n) \in C$. This is an IBIS group. It acts on $n|F|$ points, and has rank equal to the dimension of the code; if there is no coordinate in which all codewords are zero, then it has n orbits each of size $|F|$.

The classification problem for primitive IBIS groups is likely to be easier, though even that has not yet been done. In [Cameron and Fon-Der-Flaass 1995], the IBIS groups whose associated matroid is a uniform matroid are determined; these are Frobenius and Zassenhaus groups and their analogues, that is, groups which, for some positive integer t, are t-transitive and have the property that the pointwise stabiliser of any $t + 1$ points is trivial. (The *uniform* matroid of rank r is the one whose bases are all the sets of size r of the ground set.) All such finite groups with $t > 1$ (that is, those which are not Frobenius groups) have been explicitly determined (without using CFSG), by Zassenhaus, Feit, Ito and Suzuki for $t = 2$, and by Gorenstein and Hughes for larger values.

(However, infinite examples are easy to construct and exist in profusion: there is an action of the free group of countable rank with this property for any value of t.)

It is also not known whether there is a similar geometric characterisation of groups in which all greedy bases have the same size.

Blaha [1992] showed that irredundant and greedy bases are not too much larger than the smallest possible:

Theorem 2. *Let G be a permutation group of degree n with minimal base size $b(G)$. Then*

(a) *any irredundant base for G has size at most $b(G)\log n$ (logarithm to base 2);*

(b) *any greedy base for G has size at most $b(G)(\log\log n + c)$.*

Blaha proved that these bounds are essentially best possible. But for primitive groups, stronger results should be possible. It is conjectured, for example, that if G is primitive, then a greedy base for G has size at most $cb(G)$, where c is a universal constant. Indeed, the limit superior of the ratio of greedy base size to base size, as $b(G)\to\infty$, is conjectured to be $\frac{9}{8}$. The extreme examples involve the symmetric group S_m acting on the set of 2-element subsets of $\{1,\ldots,m\}$. The greedy algorithm chooses disjoint 2-sets until almost all elements of $\{1,\ldots,m\}$ have been covered, and then has to go back and extend two disjoint pairs to a 4-vertex path, giving a base of size roughly $\frac{3}{4}m$; on the other hand, covering most of $\{1,\ldots,m\}$ by 3-vertex paths gives a base of size roughly $\frac{2}{3}m$.

Recently, Coen Del Valle and Colva Roney-Dougal have given the exact value of the base size for the symmetric group of degree n acting on r-sets for $2\le r\le n/2$. The result is complicated to state, depending on the relative sizes of n and r.

We conclude with two further occurrences of bases.

(a) The first fractional exponential bound for the order of a uniprimitive (primitive but not 2-transitive) permutation group of degree n was found by Babai [1981]. He showed that such a group has a base whose size is bounded by $4\sqrt{n}\log n$. It is clear that a group with degree n and a base of size b has order at most n^b. (Soon after Babai's result appeared, it was observed that much stronger results could be found using the classification of finite simple groups: a bound $n^{c\log n}$ with "known" exceptions. These exceptions are the so-called *large-base groups* which are explained below.)

(b) Graph theorists have considered the *metric dimension* of a connected graph, the smallest d for which there is a d-tuple $(v_1,\ldots,v_d)$ of vertices such that any vertex is uniquely determined by its d-tuple of distances from these vertices. It is clear that such a d-tuple is a base for the automorphism group of the graph. The occurrence of similar concepts in very different fields led to a lot of repetition and rediscovery, which my survey with Robert Bailey [Bailey and Cameron 2011] sets out to clear up.

These two things are related. Babai's proof involved constructing from the group a set of binary relations called a *coherent configuration* and showing that this configuration has relatively small "dimension" (using the relations in the configuration in place of graph distances).

A large-base group is either a symmetric or alternating group S_n or A_n in its action on the set of k-subsets of $\{1,\ldots,n\}$, or a subgroup of the wreath product of such a group with the symmetric group of degree l containing the socle A_n^l of this group. Their base sizes are fractional powers of the degree, and so their orders are roughly $n^{n^{1/kl}}$. Often in computational group theory it is necessary to treat the large-base groups separately.

There has been a lot of very recent activity around permutation group bases. Scott Harper remarked that the result about IBIS groups gives us powerful information about permutation groups where all irredundant bases have the same size, but the groups for which all minimal bases have the same size has at present no comparable theory. One could ask similar questions about "greedy bases" in Blaha's sense.

It is also appropriate to mention here the work of Gill, Lodà and Spiga [Gill et al. 2022] on a parameter they call *height*, which is the maximum size of an independent set (where a set is *independent* if its pointwise stabiliser is properly contained in the pointwise stabiliser of any subset). They showed that the height of a primitive permutation group of degree n which is not a large-base group is smaller than $9\log n$. This parameter then gives a bound for the *relational complexity* of a permutation group, a parameter introduced by Cherlin [Cherlin 2016; Cherlin et al. 1996], in connection with the model theory of finite permutation groups: the relational complexity is at most the height plus one.

To elaborate: the relational complexity of G is the least k for which G is an automorphism group of a homogeneous relational structure with arity k; more precisely, it is the least k such that, for any $n \geq k$, if $(a_1, \dots, a_n)$ and $(b_1, \dots, b_n)$ are two n-tuples of points, they lie in the same G-orbit if and only if all corresponding sub-k-tuples $(a_{i_1}, \dots, a_{i_k})$ and $(b_{i_1}, \dots, b_{i_k})$ lie in the same G-orbit.

Gill et al. also proved a similar bound on the maximum size of an irredundant base for a primitive permutation group.

6. Finite group parameters

In this final section I discuss a few "measures" of a finite group which are related to base size of permutation actions of the group. As we will see, some of these can be defined in terms of the subgroup lattice of G, a topic with a long history but still many open problems: see [Schmidt 1994] for a fairly recent account.

The smallest number $d(G)$ of generators of G is not a good measure of size, since arbitrarily large finite groups (such as symmetric groups) are 2-generated. We can avoid this problem as follows. If f is a function from finite groups to natural numbers, let
$$f^{\uparrow}(G) = \max\{f(H) : H \leq G\}.$$
For any f, the function $f^{\uparrow}$ is monotonic (in the sense that $G \leq H$ implies $f^{\uparrow}(G) \leq f^{\uparrow}(H)$). For example, for the symmetric groups, we have $d^{\uparrow}(S_n) = \lfloor n/2 \rfloor$ for $n > 3$ [McIver and Neumann 1987].

Define two other measures as follows:

(a) $\mu(G)$ is the maximal size of a minimal (under inclusion) generating set for G. The parameter $\mu(G)$ is important in the analysis of a random walk on generating sets for G; see [Diaconis and Saloff-Coste 1998]. For the symmetric groups we have $\mu(S_n) = \mu^{\uparrow}(S_n) = n-1$ [Whiston 2000].

(b) $l(G)$ is the length of the largest subgroup chain in G. This is an interesting measure which bounds various other measures, and was considered by Babai [1986]. It has the nice properties that it is monotonic and, if N is a normal subgroup of G, then $l(G) = l(N) + l(G/N)$; so its value is determined by the composition factors of G. In 1982, I showed that

$$l(S_n) = \left\lceil \frac{3n}{2} \right\rceil - b(n) - 1,$$

where $b(n)$ is the number of 1s in the base 2 representation of n. This appears in a paper with Solomon and Turull [Cameron et al. 1989]; these authors have computed $l(G)$ for various simple groups G.

Given a finite group G, we define three numbers $b_1(G)$, $b_2(G)$, $b_3(G)$ as follows. In each case, the maximum is taken over all permutation representations of G (not necessarily faithful).

- $b_1(G)$ is the maximum, over all representations, of the maximum size of an irredundant base.
- $b_2(G)$ is the maximum, over all representations, of the maximum size of a minimal base.
- $b_3(G)$ is the maximum, over all representations, of the minimum base size.

Clearly we have $b_3(G) \le b_2(G) \le b_1(G)$. These inequalities can be strict. The group $G = \mathrm{PSL}(2, 7)$ has $b_1(G) = 5$, $b_2(G) = 4$, and $b_3(G) = 3$.

Proposition 1. $$b_1(G) = l(G).$$

Proof. An irredundant base $(x_1, \dots, x_k)$ gives a descending chain of subgroups $G = G_0 > G_1 > \cdots > G_k$, where G_i is the pointwise stabiliser of $\{x_1, \dots, x_i\}$. Conversely, given a chain of subgroups, take the union of the coset spaces of these subgroups, and form a base by choosing the given subgroups in the order given. □

There is a connection between b_2 and μ. Let $B(n)$ denote the Boolean lattice of subsets of an n-set, and $L(G)$ the subgroup lattice of G.

Proposition 2. *Let G be a finite group.*

(a) *The largest n such that $B(n)$ is embeddable as a join-semilattice of $L(G)$ is $\mu^{\uparrow}(G)$.*

(b) *The largest n such that $B(n)$ is embeddable as a meet-semilattice of $L(G)$ in such a way that the minimal element is a normal subgroup of G is $b_2(G)$.*

(c) *$B(n)$ is embeddable as a meet-semilattice in $L(G)$ if and only if it is embeddable as a join-semilattice.*

Proof. (a) If $\{g_1, \dots, g_n\}$ is an independent set in G, then the subgroups generated by subsets of this set form a join-semilattice isomorphic to $B(n)$. Conversely, given such a semilattice of the subgroup lattice, choose elements g_i contained in all the maximal subgroups except the i-th.

(b) Given a minimal base of size n, the subgroups stabilising subsets of the base form a meet-semilattice whose minimal element is the kernel of the group action. Conversely, suppose we have an embedding of $B(n)$ as meet-semilattice. Then, reversing order, we have subgroups H_I for each $I \subseteq N = \{1, \dots, n\}$, with H_N normal in G and $H_i \cap H_J = H_{I \cup J}$. Consider the permutation representation on the union of the coset spaces $H_{\{i\}}$ for $i \in N$. The kernel of this representation is H_N, and the subgroups $H_{\{i\}}$ form a minimal base of size n.

(c) Suppose first that $B(n)$ is a join-semilattice of $L(G)$. Let $N = \{1, \dots, n\}$. Then, for every subset I of N, there is a subgroup H_I of G, and $H_{I \cup J} = \langle H_I, H_J \rangle$ for any two subsets I and J. Moreover, all these subgroups are distinct. In particular, $H_i \not\leq H_{N \setminus \{i\}}$ for all i (where H_i is shorthand for $H_{\{i\}}$); else

$$H_N = \langle H_i, H_{N \setminus \{i\}} \rangle = H_{N \setminus \{i\}},$$

contrary to assumption.

Let $K_i = H_{N \setminus \{i\}}$, and, for any $I \subseteq N$, put

$$K_I = \bigcap_{i \in I} K_i,$$

with the convention that $K_\varnothing = G$. We claim that all the subgroups K_I are distinct. Suppose that two of them are equal, say $K_I = K_J$. By interchanging I and J if necessary, we may assume that there exists $i \in I \setminus J$. But then $H_i \leq K_J$ while $H_i \not\leq K_I$, a contradiction.

Now it is clear that $K_I \cap K_J = K_{I \cup J}$, so that we have an embedding of $B(n)$ as meet-semilattice (where we have reversed the order-isomorphism to simplify the notation).

The reverse implication is proved by an almost identical argument. $\square$

It is not known whether the extra condition in (b) is really necessary: perhaps $b_2(G) = \mu^{\uparrow}(G)$ for any G. (Note that $\mu^{\uparrow}(G)$ is the size of the largest independent set of elements of G.)

Much less is known about $b_3(G)$. If G is a nonabelian finite simple group, then $b_3(G)$ can be computed by looking only at the primitive actions of G.

One could ask similar questions about greedy bases. Nothing is known.

Another question: in which of these results can the use of CFSG be avoided?

Acknowledgement

I am very grateful to two reviewers whose careful and thoughtful reviews have substantially improved this paper.

References

[Araújo and Fountain 2004] J. Araújo and J. Fountain, "The origins of independence algebras", pp. 54–67 in *Semigroups and languages* (Lisbon, 2002), edited by I. M. Araújo et al., World Sci. Publ., River Edge, NJ, 2004. MR Zbl

[Araújo et al. 2011] J. Araújo, M. Edmundo, and S. Givant, "v^*-algebras, independence algebras and logic", *Internat. J. Algebra Comput.* **21**:7 (2011), 1237–1257. MR Zbl

[Araújo et al. 2022] J. Araújo, W. Bentz, P. J. Cameron, M. Kinyon, and J. Konieczny, "Matrix theory for independence algebras", *Linear Algebra Appl.* **642** (2022), 221–250. MR Zbl

[Babai 1981] L. Babai, "On the order of uniprimitive permutation groups", *Ann. of Math.* (2) **113**:3 (1981), 553–568. MR Zbl

[Babai 1986] L. Babai, "On the length of subgroup chains in the symmetric group", *Comm. Algebra* **14**:9 (1986), 1729–1736. MR Zbl

[Babai 2015] L. Babai, "Graph isomorphism in quasipolynomial time", preprint, 2015. arXiv 1512.03547

[Bailey and Cameron 2011] R. F. Bailey and P. J. Cameron, "Base size, metric dimension and other invariants of groups and graphs", *Bull. Lond. Math. Soc.* **43**:2 (2011), 209–242. MR Zbl

[Blaha 1992] K. D. Blaha, "Minimum bases for permutation groups: the greedy approximation", *J. Algorithms* **13**:2 (1992), 297–306. MR Zbl

[Cameron and Deza 1979] P. J. Cameron and M. Deza, "On permutation geometries", *J. London Math. Soc.* (2) **20**:3 (1979), 373–386. MR Zbl

[Cameron and Fon-Der-Flaass 1995] P. J. Cameron and D. G. Fon-Der-Flaass, "Bases for permutation groups and matroids", *European J. Combin.* **16**:6 (1995), 537–544. MR Zbl

[Cameron and Szabó 2000] P. J. Cameron and C. Szabó, "Independence algebras", *J. London Math. Soc.* (2) **61**:2 (2000), 321–334. MR Zbl

[Cameron et al. 1989] P. J. Cameron, R. Solomon, and A. Turull, "Chains of subgroups in symmetric groups", *J. Algebra* **127**:2 (1989), 340–352. MR Zbl

[Cherlin 2016] G. Cherlin, "On the relational complexity of a finite permutation group", *J. Algebraic Combin.* **43**:2 (2016), 339–374. MR Zbl

[Cherlin et al. 1996] G. L. Cherlin, G. A. Martin, and D. H. Saracino, "Arities of permutation groups: wreath products and k-sets", *J. Combin. Theory Ser. A* **74**:2 (1996), 249–286. MR Zbl

[Diaconis and Saloff-Coste 1998] P. Diaconis and L. Saloff-Coste, "Walks on generating sets of groups", *Invent. Math.* **134**:2 (1998), 251–299. MR Zbl

[Fountain and Lewin 1992] J. Fountain and A. Lewin, "Products of idempotent endomorphisms of an independence algebra of finite rank", *Proc. Edinburgh Math. Soc.* (2) **35**:3 (1992), 493–500. MR Zbl

[Giansiracusa and Giansiracusa 2018] J. Giansiracusa and N. Giansiracusa, "A Grassmann algebra for matroids", *Manuscripta Math.* **156**:1-2 (2018), 187–213. MR Zbl

[Gill et al. 2022] N. Gill, B. Lodà, and P. Spiga, "On the height and relational complexity of a finite permutation group", *Nagoya Math. J.* **246** (2022), 372–411. MR Zbl

[Gould 1995] V. Gould, "Independence algebras", *Algebra Universalis* **33**:3 (1995), 294–318. MR Zbl

[Grätzer 1963] G. Grätzer, "A theorem on doubly transitive permutation groups with application to universal algebras", *Fund. Math.* **53** (1963), 25–41. MR Zbl

[Kerby 1974] W. Kerby, *On infinite sharply multiply transitive groups*, Hamburger Mathematische Einzelschriften (N.F.) **6**, Vandenhoeck & Ruprecht, Göttingen, 1974. MR Zbl

[Maund 1989] T. Maund, *Bases for permutation groups*, D.Phil. thesis, Oxford University, 1989.

[McIver and Neumann 1987] A. McIver and P. M. Neumann, "Enumerating finite groups", *Quart. J. Math. Oxford Ser.* (2) **38**:152 (1987), 473–488. MR Zbl

[Rips et al. 2017] E. Rips, Y. Segev, and K. Tent, "A sharply 2-transitive group without a non-trivial abelian normal subgroup", *J. Eur. Math. Soc.* (*JEMS*) **19**:10 (2017), 2895–2910. MR Zbl

[Schmidt 1994] R. Schmidt, *Subgroup lattices of groups*, De Gruyter Expositions in Mathematics **14**, Walter de Gruyter & Co., Berlin, 1994. MR Zbl

[Sims 1970] C. C. Sims, "Computational methods in the study of permutation groups", pp. 169–183 in *Computational problems in abstract algebra* (Oxford, 1967), edited by J. Leech, Pergamon, Oxford-New York-Toronto, Ont., 1970. MR Zbl

[Tits 1952] J. Tits, "Sur les groupes doublement transitifs continus", *Comment. Math. Helv.* **26** (1952), 203–224. MR Zbl

[Urbanik 1966] K. Urbanik, "Linear independence in abstract algebras", *Colloq. Math.* **14** (1966), 233–255. MR Zbl

[Whiston 2000] J. Whiston, "Maximal independent generating sets of the symmetric group", *J. Algebra* **232**:1 (2000), 255–268. MR Zbl

[Wilke 1972] F. W. Wilke, "Pseudo-fields and doubly transitive groups", *Bull. Austral. Math. Soc.* **7** (1972), 163–168. MR Zbl

[Zassenhaus 1935] H. Zassenhaus, "Über endliche Fastkörper", *Abh. Math. Sem. Univ. Hamburg* **11**:1 (1935), 187–220. MR Zbl

[Zilber 1984] B. I. Zilber, "Strongly minimal countably categorical theories, II", *Sibirsk. Mat. Zh.* **25**:3 (1984), 71–88. MR Zbl

[Zilber 1988a] B. Zilber, "Finite homogeneous geometries", pp. 186–208 in *Proceedings of the* 6*th Easter Conference on Model Theory* (Wendisch Rietz, Germany, 1988), edited by B. Dahn and H. Wolter, Seminarberichte **98**, Humboldt Univ., Berlin, 1988. MR Zbl

[Zilber 1988b] B. I. Zilber, "Hereditarily transitive groups and quasi-Urbanik structures", *Trudy Inst. Mat.* (*Novosibirsk*) **8** (1988), 58–77. In Russian; translated in *Amer. Math. Soc. Transl. Ser.* 2 **195** (1999), 165–186. MR Zbl

Received 4 Dec 2022. Revised 10 Apr 2023.

PETER J. CAMERON:

pjc20@st-andrews.ac.uk
School of Mathematics and Statistics, University of St Andrews, North Haugh, St Andrews, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.433

On the model theory of open generalized polygons

Anna-Maria Ammer and Katrin Tent

We show that for any $n \geq 3$ the theory of open generalized n-gons is complete, decidable and strictly stable, yielding a new class of examples in the zoo of stable theories.

1. Introduction

Generalized polygons were introduced by Tits in order to give geometric interpretations of the groups of Lie type in rank 2, in the same way that projective planes correspond to groups of type A_2. In fact, generalized polygons are the rank 2 case of spherical buildings. A generalized n-gon is a bipartite graph with diameter n (i.e., any two vertices have distance at most n), girth $2n$ (i.e., the smallest cycles have length $2n$) and such that all vertices have valency at least 3. Clearly, for $n = 2$ such a graph is simply a complete bipartite graph and in what follows we always assume $n \geq 3$. Thinking of the bipartition as corresponding to points and lines, we see that the case $n = 3$ is simply a different way of phrasing the axioms of a projective plane, namely, any two points lie on a unique line, any two lines intersect in a unique point and every line contains at least three points. (It then easily follows that every point has at least three lines passing through it.) Remarkably, if the graph is finite, then by a fundamental result of Feit and Higman [1964] the only possible values for n are 3, 4, 6 and 8. Similar restrictions hold for other well-behaved or *tame* categories of generalized polygons, e.g., if one assumes that the underlying sets of vertices are compact, or algebraic, one obtains the same restrictions. Since we tend to think of finite Morley rank as a rather strong tameness assumption it might be remarkable that this restriction does not hold in this setting; see [Tent 2000].

In fact, it is easy to see that infinite generalized n-gons exist for any $n \geq 3$: starting with a finite bipartite configuration that does not contain any $2k$-cycles for $k < n$, one can easily complete this by freely adding enough paths in order to make sure

The authors were partially supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure and SFB 1442.
MSC2020: 03C45, 03C65, 05C63, 51E24.
Keywords: generalized polygons, stable structures, Hrushovski limits, combinatorial geometries.

that the graph has diameter n (see Definition 2.4 below). In fact, such constructions yield the only known examples of generalized n-gons for $n \neq 3, 4, 6, 8$.

Free projective planes were studied by M. Hall [1943], Siebenmann [1965], and Kopeikina [1945] and their model theory was studied in [Tent 2011; Hyttinen and Paolini 2021; Tent and Zilber 2015]. The theory of the free projective planes is strictly stable by [Hyttinen and Paolini 2021] and the notion of independence in the sense of stability agrees with the one studied in [Tent 2014; Müller and Tent 2019]. In this note we extend the results from [Tent 2011; Hyttinen and Paolini 2021] to open generalized polygons, using the methods developed in [Tent 2000; 2014; Müller and Tent 2019]. In particular, it was shown in [Hyttinen and Paolini 2021] that the theory of open projective planes is complete, strictly stable, does not have a prime model and has uncountably many nonisomorphic countable models.

2. Generalized polygons

We first recall some graph-theoretic notions. For a and b in A, the *distance* $d(a, b)$ between a and b is the smallest number m for which there is a path

$$a = a_0, a_1, \ldots, a_m = b$$

with a_i in A, where a_i and a_{i+1} are incident for $0 \leq i < m$. We may write $d_A(a, b)$ to emphasize the dependence on the graph A.

The *girth* of a graph A is the length of a shortest cycle in A. The *diameter* of a graph A is the maximal distance between two elements in A. We say that a subgraph A of a graph B is *isometrically embedded* into B if for all $a, b \in A$ we have $d_A(a, b) = d_B(a, b)$. For a vertex $a \in A$ we write $D_1(a)$ for the set of neighbours of a. Then $|D_1(a)|$ is called the valency of a in A.

From now on we fix $n \geq 3$.

Definition 2.1. A *weak generalized n-gon* Γ is a bipartite graph with diameter n and girth $2n$. If Γ is *thick*, i.e., if each vertex has valency at least 3, then Γ is a *generalized* n-gon.

A *partial* n-gon is a connected bipartite graph of girth at least $2n$.

A (partial) n-gon Γ_0 is *nondegenerate* if Γ_0 contains a cycle of length at least $2n + 2$ or a path $\gamma = (x_0, \ldots, x_{n+3})$ with $d_{\Gamma_0}(x_0, x_{n+3}) = n + 3$.

A (generalized) n-gon Γ_0 contained as a subgraph in a generalized n-gon Γ is called a (generalized) sub-n-gon of Γ.

Remark 2.2. Note that every thick generalized n-gon is nondegenerate.

The assumption that a partial n-gon is connected is not strictly necessary (and it is not required in [Hall 1943] for $n = 3$). Note that for $n = 3$ any two distinct points have distance 2 (and similarly for lines). This is not true anymore for $n > 3$, so the requirement that the graph is connected prevents ambiguities.

Definition 2.3. Let $(x = x_0, \dots, x_{n-1} = y)$ be a path in Γ. If every x_i, $1 \le i \le n-2$, has valency 2 in Γ, then $(x_1, \dots, x_{n-2})$ is called a clean arc in Γ (with endpoints x, y). A *loose end* is a vertex of valency at most 1.

A *hat-rack* of length $k \ge n+3$ is a path $(x_0, \dots, x_k)$ together with subsets of $D_1(x_i)$ for $1 \le i \le k-1$.

The following definition is due to Tits [1977], who first introduced free extensions for generalized polygons, expanding earlier definitions by M. Hall and Siebenmann [Hall 1943; Siebenmann 1965].

Definition 2.4. Let Γ_0 be a partial n-gon. We define the free completion of Γ_0 by induction on $i < \omega$ as follows:

For $i \ge 0$ we obtain Γ_{i+1} from Γ_i by adding a clean arc between every two elements of Γ_i which have distance $n+1$ in Γ_i. Then $\Gamma = F(\Gamma_0) = \bigcup_{i<\omega} \Gamma_i$ is called the *free n-completion* of Γ_0. We say that Γ is *freely generated* over Γ_0.

Note that if Γ_0 does not contain vertices at distance $\ge n+1$, then $F(\Gamma_0) = \Gamma_0$. Also note that by adding a clean arc between vertices of distance $n+1$ we are creating a new cycle of length $2n$.

Remark 2.5. If two elements in a generalized n-gon Γ have distance less than n, there is a unique shortest path in Γ connecting them (otherwise we would obtain a short cycle).

A weak generalized n-gon which contains a $2(n+1)$-cycle is a generalized n-gon (see [van Maldeghem 1998, Section 1.3]). Hence if Γ_0 is a partial, nondegenerate n-gon, then $F(\Gamma_0) = \Gamma$ contains a $2(n+1)$-cycle and in fact, Γ is an infinite generalized n-gon [van Maldeghem 1998] and every vertex z in $F(\Gamma_0)$ has infinite valency.

We also note the following for future reference:

Remark 2.6. Let Γ be a generalized n-gon and let $\gamma \subset \Gamma$ be a $2n$-cycle. Then for any $x \in \gamma$ there is a unique $x' \in \gamma$ with $d(x, x') = n$ (x' is called the opposite of x in γ), and for any $y \in D_1(x) \setminus \gamma$ there is a unique $y' \in D_1(x')$ such that $d(y, y') = n-2$. Note that the result of adding a clean arc to $\gamma \cup \{y\}$ is the same as adding a clean arc to $\gamma \cup \{y'\}$.

Definition 2.7. Let Γ be a generalized n-gon and $A \subset \Gamma$. Then $\langle A \rangle_\Gamma$ denotes the intersection of all generalized sub-n-gons of Γ containing A. For $\Gamma_0 \subset \Gamma$ we put $\langle A \rangle_{\Gamma_0} = \langle A \rangle_\Gamma \cap \Gamma_0$.

Remark 2.8. If $A \subseteq \Gamma_0 \subseteq \Gamma$, then $\langle A \rangle_{\Gamma_0}$ is the intersection of all $B \supset A$, $B \subseteq \Gamma_0$, such that A is isometrically contained in B. If A is nondegenerate, then $\langle A \rangle_\Gamma$ is a generalized sub-n-gon of Γ, the n-gon (not necessarily freely) generated by A in Γ. Since shortest paths between vertices at distance $n-1$ are unique, clearly $\langle A \rangle_\Gamma \subseteq \operatorname{acl}(A)_\Gamma$. If $\Gamma = F(A)$, then $\langle A \rangle_\Gamma = \Gamma$.

We note the following useful observations:

Lemma 2.9. *Let Γ_0 be a nondegenerate partial n-gon, and let $\Gamma = F(\Gamma_0) = \bigcup \Gamma_i$ be as in Definition 2.4.*

(i) *If $A \subset \Gamma_k \setminus \Gamma_i$ is isometrically embedded into Γ_k, then $\langle A \rangle_\Gamma$ does not intersect Γ_i and $\langle A \rangle_\Gamma = F(A)$.*

(ii) *If $A \subset \Gamma_0$ is such that $\Gamma_k \setminus A$ is isometrically embedded into Γ_k, then $\langle \Gamma_k \setminus A \rangle_\Gamma$ does not intersect A.*

(iii) *Any automorphism of Γ_0 extends to an automorphism of Γ.*

Proof. All parts follow directly from the construction: e.g., for (i) it suffices to show inductively that $\Gamma_1(A)$ is isometrically embedded into $\Gamma_{k+1} \setminus \Gamma_i$. Then (i) follows by induction. Let $\gamma \subset \Gamma_{k+1} \setminus \Gamma_i$ be a clean arc connecting $a, b \in A$ with $d_A(a, b) = n + 1$. Any $c \in \gamma$ has valency 2, so any path from c to an element in A passes through a or b. Since A is isometrically embedded in Γ_k, the claim follows. The proof for part (ii) is similar and part (iii) is clear. $\square$

Now we can state the main definition of this note, extending the definition of free and open projective planes from [Hall 1943] to generalized n-gons.

Definition 2.10. A (partial) generalized n-gon Γ is *open* if every finite subgraph contains a loose end or a clean arc.

We call a generalized n-gon Γ *free* if it is the free n-completion of a hat-rack of length at least $n + 3$. In particular, we let Γ^k denote the free n-completion of the path $\gamma_k = (x_0, \ldots, x_k)$ for $k \geq n + 3$.

Note that Γ^k is a free generalized n-gon for $k \geq n + 3$.

Remark 2.11. Clearly, every free generalized n-gon is open. Beware, however, that the converse is not true in general (see Proposition 3.16), but holds for finitely generated generalized n-gons (see Proposition 4.1).

Clearly, as observed by [Hyttinen and Paolini 2021] for the case $n = 3$ being an open generalized n-gon is a first-order property. We can therefore define:

Definition 2.12. Let T_n denote the theory of open generalized n-gons in the language of graphs expanded by predicates for the bipartition.

Note that T_n is $\forall\exists$-axiomatizable. We start with some easy observations:

Remark 2.13. It follows immediately from Remark 2.5 and the definition of an open generalized n-gon that for $M \models T_n$ and a nondegenerate subgraph $A \subseteq M$ we have $\operatorname{acl}(A) \models T_n$. In other words, every algebraically closed nondegenerate subset of a model of T_n is itself a model of T_n. Clearly, $\operatorname{acl}(A)$ is prime over A [Tent and Ziegler 2012, Section 5.3].

Remark 2.14. Let $T_{n,\gamma}$ be the theory T_n expanded by constants for the vertices of a path $\gamma = (a_0, \dots, a_{n+3})$. Then $F(\gamma)$ is the prime model of $T_{n,\gamma}$ since $F(\gamma)$ is algebraic over γ, hence countable and atomic, hence prime (see [Tent and Ziegler 2012, Theorem 4.5.2]).

This is similar to the situation in free groups described in Sela's seminal results, but obviously much easier to prove in the current setting: both theories are strictly stable, and only the "natural embeddings" are elementary. Namely, we will see later that $\Gamma^k \preccurlyeq \Gamma^m$ if and only if $k \leq m$ and the embedding is the natural one.

Adapting[1] Siebenmann's definition for the case $n = 3$ [Siebenmann 1965] we define:

Definition 2.15. If A is a partial n-gon, a *hyperfree minimal extension* of A is an extension by a clean arc between two elements $a, b \in A$ with $d_A(a, b) = n + 1$ or by a loose end.

Let Γ and Γ' be partial n-gons. We say that Γ is *HF-constructible* from Γ' (or over Γ') if there is an ordinal α and a sequence $(\Gamma_\beta)_{\beta<\alpha}$ of partial n-gons such that

(i) $\Gamma_0 = \Gamma'$;

(ii) if $\beta = \gamma + 1$, then Γ_β is a hyperfree minimal extension of Γ_γ;

(iii) if β is a limit ordinal, then $\Gamma_\beta = \bigcup_{\gamma<\beta} \Gamma_\gamma$;

(iv) $\Gamma = \bigcup_{\beta<\alpha} \Gamma_\beta$.

Clearly, any free completion of a partial n-gon Γ_0 is HF-constructible from Γ_0.

As in [Hyttinen and Paolini 2021] one can show that any open generalized n-gon has an HF-ordering, but since we will not be using this ordering, we omit the details.

Definition 2.16. Let $A, B \subseteq M \models T_n$, $A \cap B = \varnothing$. We call B *closed over* A if B contains neither a clean arc with endpoints in $A \cup B$ nor a loose end. We say that B is *open over* A if B contains no finite set closed over A and in this case we write $A \leq_o A \cup B$. We write $\widehat{A}_M = A \cup \bigcup\{B_0 \subset M \mid B_0 \text{ finite and closed over } A\}$.

Remark 2.17. Note that if B_1, B_2 are closed over A, then so is $B_1 \cup B_2$.

Lemma 2.18. *If B is open over A and $B \setminus A$ is finite, then B is HF-constructible over A. In particular, if A is a finite open partial n-gon, then A is HF-constructible from the empty set. Moreover, if $A \leq_o \Gamma$, where Γ is a generalized n-gon, then $F(A) \cong \langle A\rangle_\Gamma \subseteq \Gamma$.*

Proof. If $B \setminus A$ is a minimal counterexample, then it cannot contain either a loose end or a clean arc, contradicting the assumption of B being open over A. $\square$

[1] Note that Siebenmann also allows adding vertices of valency 0.

Now consider the class $\mathcal{K}$ of finite open partial n-gons (in the language of bipartite graphs) with strong embeddings given by $\leq_o$. Note that $\mathcal{K}$ is contained both in the class of partial n-gons considered in [Tent 2011] as well as in the class of partial n-gons considered in [Tent 2000] (see Lemma 3.12).

Definition 2.19. For graphs $A \subseteq B, C$, let $B \otimes_A C$ denote the free amalgam of B and C over A.

Let $A \leq_o B, C$ be open partial n-gons (contained in some generalized n-gon Γ) with $\langle A\rangle_B = \langle A\rangle_C = A$. Then we call $B \oplus_A C := F(B \otimes_A C)$ the *canonical amalgam* of B and C over A.

The canonical amalgam was used in [Tent 2011] (and for $n = 3$ in [Hyttinen and Paolini 2021]).

Remark 2.20. (i) If $A \leq_o B, C$ are open partial n-gons with $\langle A\rangle_B = \langle A\rangle_C = A$, then $B, C \leq_o B \otimes_A C \leq_o B \oplus_A C$. If $B \otimes_A C$ is nondegenerate, then $B \oplus_A C$ is an open generalized n-gon.

(ii) If $B \cap C = A$ and $B \cup C \leq_o \Gamma$ for some generalized n-gon Γ, then $B \cup C \cong B \otimes_A C$ and $\langle B \cup C\rangle_\Gamma \cong B \oplus_A C$.

The following is as in [Tent 2000; 2011; Hyttinen and Paolini 2021]:

Proposition 2.21. *Let $\mathcal{K}$ be the class of finite connected open partial n-gons. Then $(\mathcal{K}, \leq_o)$ satisfies*

(i) *amalgamation: if $A, B_1, B_2 \in \mathcal{K}$ such that $\iota_i : A \to B_i$ and $\iota_i(A) \leq_o B_i$, $i = 1, 2$, then there exist $C \in \mathcal{K}$ and $\kappa_i : B_i \to C$, $i = 1, 2$ such that $\kappa_i(B_i) \leq_o C$, $i = 1, 2$, and $\kappa_1(\iota_1(a)) = \kappa_2(\iota_2(a))$ for all $a \in A$.*

(ii) *joint embedding: for any two graphs $A, B \in \mathcal{K}$ there is some $C \in \mathcal{K}$ such that A, B can be strongly embedded (in the sense of $\leq_o$) into C.*

Hence the limit $\Gamma_{\mathcal{K}}$ exists and is an open generalized n-gon.

Proof. Since $\varnothing \in \mathcal{K}$, it suffices to verify the amalgamation property. Inductively we may assume that B is a minimal hyperfree extension of A, so either a clean arc or a loose end. If C does not contain a copy of B over A, then $B \otimes_A C \in \mathcal{K}$ and this is enough. $\square$

Note that the class $(\mathcal{K}, \leq_o)$ is unbounded in the sense that for any $A \in \mathcal{K}$ there exists some $B \in \mathcal{K}$ with $A \neq B$ and $A \leq_o B$.

Definition 2.22. Let $M \models T_n$. Then we say that M is $\mathcal{K}$-saturated if for all finite sets $A, B \in \mathcal{K}$ with $A \leq_o B$ and any copy A' of A strongly embedded into M there is a strong embedding of B over A' into M.

Note that by construction, $\Gamma_{\mathcal{K}}$ is $\mathcal{K}$-saturated and that (as in any such Hrushovski construction) every $\mathcal{K}$-saturated structure is $\mathcal{K}$-homogeneous in the sense that any partial automorphism between strongly embedded substructures extends to an automorphism.

Theorem 2.23. *For any $n \geq 3$, the theory T_n of open generalized n-gons is complete and hence decidable.*

Proof. Let $M \models T_n$. It suffices to show that M is elementarily equivalent to $\Gamma_{\mathcal{K}}$. Clearly we may assume that M is ω-saturated and we claim that any ω-saturated M is $\mathcal{K}$-saturated: Let $A \leq_o B$ be from $\mathcal{K}$ and assume that $A \leq_o M$ (via some strong embedding). We have to show that we can find an embedding B' of B into M such that there does not exist a finite set closed over B' in M. This is clear if B is an extension of A by a clean arc since such paths are unique. If B is an extension of A by a loose end b, then the type of b over A expressing that there is no finite set D closed over $A \cup \{b\}$ is realized in $\Gamma_{\mathcal{K}}$, so it is consistent and therefore realized in M by ω-saturation. Now both M and $\Gamma_{\mathcal{K}}$ are $\mathcal{K}$-saturated from which it follows (by standard back-and-forth) that they are partially isomorphic and hence elementarily equivalent. □

We say that a set B neighbours a set A if every $a \in A$ has a neighbour in $B \setminus A$.

Lemma 2.24. *Let $M \models T_n$, $A \subset M$ finite. Then M does not contain three disjoint sets B_1, B_2, B_3 each closed over A and neighbouring A. In particular, if B is closed over A, then B is algebraic over A.*

Proof. Consider $C = A \cup B_1 \cup B_2 \cup B_3 \subset M$. Then every vertex in A has valency at least 3 in C and C contains no clean arc. It follows that C is not open, contradicting $M \models T_n$.

Now suppose B is minimally closed over A and not algebraic over A with $|B \setminus A|$ minimal. Since B is not algebraic over A, we find disjoint copies B_1, B_2, B_3 of B over A, contradicting the first part of the lemma. □

Lemma 2.24 directly implies:

Corollary 2.25. *If $M \models T_n$ and $A \subseteq M$, then $\operatorname{acl}(A)_M \leq_o M$.*

Definition 2.26. Let $B \models T_n$ and let A be a subgraph of B. We put

(i) $\mathrm{Cl}_0(A)_B = A$;

(ii) $\mathrm{Cl}_{i+1}(A)_B = \langle \widehat{\mathrm{Cl}(A)_i} \rangle_B$ (see Definition 2.16);

(iii) $\mathrm{Cl}_B(A) = \bigcup_{i<\omega} \mathrm{Cl}_i(A)_B$.

In other words, $\mathrm{Cl}(A)_B$ is the limit obtained from alternating between adding all closed finite subsets, and completing the partial n-gons in B.

Remark 2.27. For any subset A of $B \models T_n$ we have $\mathrm{Cl}_B(A) \leq_o B$ and by Lemma 2.24 $\mathrm{Cl}_B(A) \subseteq \mathrm{acl}_B(A)$.

Theorem 2.28. *Let $A, B \models T_n$ and $A \subseteq B$. The following are equivalent*:

(i) $A = \mathrm{acl}_B(A)$;

(ii) $A = \mathrm{Cl}_B(A)$;

(iii) $A \leq_o B$;

(iv) $A \preccurlyeq B$.

Proof. (i) implies (ii): This follows from Lemma 2.24.

(ii) implies (iii): This is by Remark 2.27.

(iii) implies (iv): By taking appropriate elementary extensions we may assume that A, B are ω_0-saturated and hence $\mathcal{K}$-saturated by the proof of Theorem 2.23. We use Tarski's test: Let $B \models \exists x\, \varphi(x, \bar{a})$ for some tuple $\bar{a} \subset A$ and let $b \in B$ such that $B \models \varphi(b, \bar{a})$. We find a countable set A_0 containing $\bar{a}$ such that $A_0 \leq_o A$ and similarly we find a countable set B_0 containing $A_0 \cup \{b\}$ such that $A_0 \leq_o B_0 \leq_o B$. Thus by $\mathcal{K}$-saturation we can embed B_0 over A_0 into A.

(iv) implies (i): This is also proved by Tarski's test. □

Corollary 2.29. *For $n+3 \leq k \leq m \leq \omega$ we have $\Gamma_k \preccurlyeq \Gamma_m$, i.e., the free generalized n-gons Γ^k form an elementary chain.*

The following lemma will be used in the proof of Theorem 2.32:

Lemma 2.30. *Let $M \models T_n$ and $A, C \subseteq M$, A finite and C algebraically closed. Then there exist $a \in A$ and $B_A = \{b_1, b_2\} \subset D_1(a)$ such that for any set B closed over $C \cup A$ and neighbouring A we have $B \cap B_A \neq \varnothing$.*

Proof. Suppose otherwise. Then by Remark 2.17 there is a set B closed over $C \cup A$ and neighbouring A such that for all $a \in A$ we have $|B \cap D_1(a)| \geq 3$. Since C is algebraically closed, we know that $B \cup A$ is open over C, so contains a loose end or a clean arc which is impossible since all $a \in A$ have valency at least 3 in $B \cup A$. □

Note that $B_A \subset \mathrm{acl}(AC)$ and B_A might be a singleton.

Exactly as in [Tent 2014] and [Müller and Tent 2019] we now define the following notion of independence (see also [Hyttinen and Paolini 2021]).

Definition 2.31. For any subsets A, B, C of the monster model $\mathfrak{M}$ of T_n, we call B and C independent over A, written $B \mathop{\smile\!\!\!\!\!\!|}^{*}_{A} C$, if

$$\mathrm{acl}(ABC) = \mathrm{acl}(AB) \oplus_{\mathrm{acl}(A)} \mathrm{acl}(AC).$$

Note that $B \mathop{\smile\!\!\!\!\!\!|}^{*}_{A} C$ implies $\mathrm{acl}(BA) \cup \mathrm{acl}(AC) \cong \mathrm{acl}(BA) \otimes_A \mathrm{acl}(CA)$.

We show that T_n is stable by establishing that $\mathop{\smile\!\!\!\!\!\!\mid}^*$ satisfies the required properties of forking as in [Tent and Ziegler 2012, Theorem 8.5.10], where in the notation of that theorem, $B \mathop{\smile\!\!\!\!\!\!\mid}^*_A C$ should be read as $\mathrm{tp}(A/C) \sqsubseteq \mathrm{tp}(A/BC)$.

Theorem 2.32. *The theory T_n of open generalized n-gons is stable. In T_n, the notion $\mathop{\smile\!\!\!\!\!\!\mid}^*$ satisfies the properties of stable forking*:

- *invariance*: *$\mathop{\smile\!\!\!\!\!\!\mid}^*$ is invariant under* $\mathrm{Aut}(\mathfrak{M})$.
- *local character*: *For all $A \subseteq \mathfrak{M}$ finite and $C \subseteq \mathfrak{M}$ arbitrary, there is some countable set $C_0 \subseteq C$ such that $A \mathop{\smile\!\!\!\!\!\!\mid}^*_{C_0} C$.*
- *weak boundedness*: *For all $B \subseteq \mathfrak{M}$ finite and $A \subseteq \mathfrak{M}$ arbitrary, there is some cardinal μ such that there are at most μ isomorphism types of $B' \subseteq \mathfrak{M}$ over C where $B' \cong_A B$ and $B' \mathop{\smile\!\!\!\!\!\!\mid}^*_A C$.*
- *existence*: *For all $B \subseteq \mathfrak{M}$ finite and $A \subseteq C \subseteq \mathfrak{M}$ arbitrary, there is some B' such that* $\mathrm{tp}(B/A) = \mathrm{tp}(B'/A)$ *and $B' \mathop{\smile\!\!\!\!\!\!\mid}^*_A C$.*
- *transitivity*: *If $B \mathop{\smile\!\!\!\!\!\!\mid}^*_A C$ and $B \mathop{\smile\!\!\!\!\!\!\mid}^*_{AC} D$ then $B \mathop{\smile\!\!\!\!\!\!\mid}^*_A CD$.*
- *weak monotonicity*: *If $B \mathop{\smile\!\!\!\!\!\!\mid}^*_A CD$, then $B \mathop{\smile\!\!\!\!\!\!\mid}^*_A C$.*

Proof. Invariance: Clearly $\mathop{\smile\!\!\!\!\!\!\mid}^*$ is invariant under $\mathrm{Aut}(\mathfrak{M})$.

Local character: Let $A \subset \mathfrak{M}$ be finite and $C \subseteq \mathfrak{M}$ arbitrary. We construct a countable set $C_\infty \subset C$ such that $\mathrm{acl}(A \cup C_\infty) \cup C \leq_o \mathfrak{M}$. Then $B = \mathrm{acl}(A \cup C_\infty)$ is countable and by Remark 2.20(ii) we have $A \mathop{\smile\!\!\!\!\!\!\mid}^*_B C$. By Lemma 2.30 there is a finite set B_A which intersects any set B closed over $A \cup C$ and neighbouring A. Let $C_A \subset C$ be finite such that $B_A \subset \mathrm{acl}(A \cup C_A)$ and put $C_0 = C_A$, $B_0 = \mathrm{acl}(A \cup C_0)$. Suppose inductively that B_i, C_i have been defined, where B_i, C_i are countable. For a finite subset $X \subset B_i$ let B_X be the finite set intersecting any set D closed over $X \cup C$ and neighbouring X, and let $C_X \subset C$ be finite such that $B_X \subset \mathrm{acl}(C_i \cup C_X)$. Put $C_{i+1} = C_i \cup \bigcup\{C_X \mid X \subset B_i \text{ finite}\}$ and $B_{i+1} = \mathrm{acl}(A \cup C_{i+1})$. Note that C_{i+1}, B_{i+1} are again countable. Finally put $C_\infty = \bigcup_{i<\omega} C_i$.

We now claim that $\mathrm{acl}(A \cup C_\infty) \cup C \leq_o \mathfrak{M}$. Suppose otherwise and let D be a finite set closed over $\mathrm{acl}(A \cup C_\infty) \cup C$ (in particular, by the definition of being closed, $D \cap (\mathrm{acl}(A \cup C_\infty) \cup C) = \varnothing$). Let Z be the set of neighbours of D in $\mathrm{acl}(A \cup C_\infty)$. Since any element of D has at most one neighbour in $\mathrm{acl}(A \cup C_\infty)$ by Theorem 2.28, we have $|Z| \leq |D|$ and hence $Z \subseteq B_i$ for some $i < \omega$. Then by construction D is closed over $Z \cup C \subseteq \mathrm{acl}(A \cup C_\infty) \cup C$ and neighbours Z, so intersects the set B_Z nontrivially. But $B_Z \subset B_{i+1}$ by construction. Since D intersects B_Z nontrivially, this contradicts our assumption $D \cap (\mathrm{acl}(A \cup C_\infty) \cup C) = \varnothing$.

Weak boundedness: Let $B \subseteq \mathfrak{M}$ be finite and $A \subseteq C \subseteq \mathfrak{M}$ be arbitrary. If $B \subset \mathrm{acl}(A)$, the claim is obvious. So assume A, C are algebraically closed and $\mathrm{tp}(B_1/A) = \mathrm{tp}(B_2/A) = \mathrm{tp}(B/A)$, so $\mathrm{acl}(B_1 A) \cong \mathrm{acl}(B_2 A)$ and B_1 and B_2 are isometric over A.

Hence from $B_1, B_2 \mathop{\smile\!\!\!\!\!\!|}^{*}_{A} C$, we have

$$\operatorname{acl}(B_1AC) \cong \operatorname{acl}(B_1A) \oplus_A C \cong \operatorname{acl}(B_2A) \oplus_A C \cong \operatorname{acl}(B_2AC).$$

In particular we have $B_1C \cong B_1 \otimes_A C \cong B_2C \cong B_2A \otimes_A C$ and so B_1 and B_2 are isometric over C. This isometry extends to an isometry from $\operatorname{acl}(B_1AC)$ to $\operatorname{acl}(B_2AC)$ fixing C and since $\operatorname{acl}(B_1AC)$ and $\operatorname{acl}(B_2C)$ are elementary substructures, this extends to an automorphism of $\mathfrak{M}$. Hence $\operatorname{tp}(B_1/C) = \operatorname{tp}(B_2/C)$.

<u>Existence</u>: Let $B \subseteq \mathfrak{M}$ be finite, $A \subset C \subseteq \mathfrak{M}$ arbitrary and $D = \operatorname{acl}(BA) \oplus_{\operatorname{acl}(A)} \operatorname{acl}(C)$. We may assume that C is nondegenerate and algebraically closed so that $C \preccurlyeq \mathfrak{M}$ and $C \preccurlyeq D$ by Theorem 2.28. By saturation and homogeneity we can embed D over C into $\mathfrak{M}$ in such a way that the image of D is an elementary substructure of $\mathfrak{M}$. Hence we find B' with $\operatorname{tp}(B/A) = \operatorname{tp}(B'/A)$ and $B' \mathop{\smile\!\!\!\!\!\!|}^{*}_{A} C$.

<u>Transitivity</u>: Let $B \mathop{\smile\!\!\!\!\!\!|}^{*}_{A} C$ and $B \mathop{\smile\!\!\!\!\!\!|}^{*}_{AC} D$, so

$$\operatorname{acl}(ABC) = \operatorname{acl}(AB) \oplus_{\operatorname{acl}(A)} \operatorname{acl}(AC)$$

and

$$\begin{aligned}\operatorname{acl}(ABCD) &\cong \operatorname{acl}(ABC) \oplus_{\operatorname{acl}(AC)} \operatorname{acl}(ACD)\\ &\cong (\operatorname{acl}(AB) \oplus_{\operatorname{acl}(A)} \operatorname{acl}(AC)) \oplus_{\operatorname{acl}(AC)} \operatorname{acl}(ACD)\\ &\cong \operatorname{acl}(AB) \oplus_{\operatorname{acl}(A)} \operatorname{acl}(ACD),\end{aligned}$$

so $B \mathop{\smile\!\!\!\!\!\!|}^{*}_{A} CD$.

<u>Weak monotonicity</u>: Let $B \mathop{\smile\!\!\!\!\!\!|}^{*}_{A} CD$, so that

$$\Gamma = \operatorname{acl}(ABCD) \cong \operatorname{acl}(AB) \oplus_{\operatorname{acl}(A)} \operatorname{acl}(ACD).$$

Now $\operatorname{acl}(AB) \otimes_{\operatorname{acl}(A)} \operatorname{acl}(AC)$ embeds isometrically into $\operatorname{acl}(AB) \otimes_{\operatorname{acl}(A)} \operatorname{acl}(ACD)$ and hence by Lemma 2.9 we have

$$\begin{aligned}\langle \operatorname{acl}(AB) \otimes_{\operatorname{acl}(A)} \operatorname{acl}(AC)\rangle_{\Gamma} &= F(\operatorname{acl}(AB) \otimes_{\operatorname{acl}(A)} \operatorname{acl}(AC))\\ &= \operatorname{acl}(AB) \oplus_{\operatorname{acl}(A)} \operatorname{acl}(AC).\end{aligned}$$ □

As a corollary of the proof we obtain:

Theorem 2.33. *The theory T of open generalized n-gons is not superstable.*

Proof. It suffices to give an example of a finite set A and an algebraically closed set C such that there is no finite set $C_0 \subset C$ with $A \mathop{\smile\!\!\!\!\!\!|}^{*}_{C_0} C$. Let

$$\Gamma_0 = \gamma_{n+3} = (x_0, \ldots, x_{n+3}) \leq_o \mathfrak{M}.$$

Then $\langle \Gamma_0 \rangle_{\mathfrak{M}} = \Gamma = \bigcup \Gamma_i$ is the free completion of Γ_0. For each $0 < i < \omega$ let $y_i \in \Gamma_i$ with $d(y_i, \Gamma_{i-1}) \geq \frac{n}{2} - 1$. Let $z_0 \neq x_{n+2}$ be a neighbour of x_{n+3} with $z_0 \mathop{\smile\!\!\!\!\!\!|}^{*}_{x_{n+2}} \Gamma_0$ and let z_i, $0 < i < \omega$, be a neighbour of y_i with $z_i \mathop{\smile\!\!\!\!\!\!|}^{*}_{y_i} \Gamma_0 z_0 \ldots z_{i-1}$. Finally connect z_i and z_{i-1} by a path λ_i of length $\geq n-1$ (depending on the parity).

Note that the resulting graph $\widetilde{\Gamma} = \Gamma \cup \bigcup_{i<\omega} \lambda_i$ is open with $\Gamma_0 \leq_o \widetilde{\Gamma}$, and hence we may assume $\widetilde{\Gamma} \leq_o \mathfrak{M}$.

Now put $A = \Gamma_0$ and $C = \operatorname{acl}(\{\lambda_i : i < \omega\})$. Then by construction there is no finite subset $C_0 \subset C$ such that $A \mathop{\smile\!\!\!\!\!\!\;|}^{*}_{C_0} C$. □

As in [Hyttinen and Paolini 2021] we can show that independence is not stationary:

Proposition 2.34. *In T_n we have* $\operatorname{acl} \neq \operatorname{dcl}$.

Proof. Let $\mathfrak{M}$ be an ω-saturated model of T_n.

If n is odd, let $A = (x_0, \dots, x_{2n+2} = x_0) \leq_o \mathfrak{M}$ be an ordered $(2n+2)$-cycle in $\mathfrak{M}$. For $i = 0, \dots, n$ let γ_i be the clean arc from x_i to x_{i+n+1} and let m_i denote the midpoint of γ_i. Let $C = A \cup \bigcup_{i=0,\dots,n} \gamma_i$. By $\mathcal{K}$-homogeneity there is an automorphism of $\mathfrak{M}$ taking A to the ordered $(2n+2)$-cycle

$$A' = (x_{n+1}, \dots, x_{2n+2} = x_0, \dots, x_{n+1}).$$

This leaves the paths γ_i, $0 \leq i \leq n$, invariant and hence fixes each m_i. This shows that $A \not\subseteq \operatorname{dcl}(m_0, \dots, m_n)$. On the other hand, C is closed over $\{m_i \mid i = 0, \dots, n\}$ and hence $A \subset \operatorname{acl}(m_0, \dots, m_{n-1})$ by Lemma 2.24.

If n is even, let $A = (x_0, \dots, x_{2n} = x_0)$ be an ordered $2n$-cycle and for $i = 1, \dots, n$ let $y_i \notin \{x_{i-1}, x_{i+1}\}$ be a neighbour of x_i and let z_i be the neighbour of x_{i+n} with $d(z_i, y_i) = n-2$. Let γ_i be the (unique) path of length $n-1$ from y_i to x_{i+n} and let m_i be its middle vertex. Note that $z_i \in \Gamma_i$. Let $D = A \cup \{y_1, \dots, y_n\}$ and assume that $D \leq_o \mathfrak{M}$. Put $C = D \cup \bigcup_{i=1,\dots,n} \gamma_i$. Then also $D' = A \cup \{z_1, \dots, z_n\} \leq_o M$. By $\mathcal{K}$-homogeneity there is an automorphism of $\mathfrak{M}$ taking D to D'. This automorphism clearly leaves A and C invariant and fixes $m_1, \dots, m_n$ pointwise, but does not fix any vertex in A. Thus as before we see that $A \not\subseteq \operatorname{dcl}(m_1, \dots, m_n)$. Since C is closed over $\{m_i \mid i = 1, \dots, n\}$ we have $A \subseteq C \subseteq \operatorname{acl}(m_1, \dots, m_n)$ by Lemma 2.24. □

3. Elementary substructures

As noted in [Tent 2011, Section 2.2], if Γ_0 is isomorphic to Δ_0, their free n-completions are also isomorphic. The reverse is obviously not true: in a completion sequence, Γ_1 and Γ_0 are not isomorphic, but they clearly have the same free n-completion.

There is nevertheless a necessary criterion for the free n-completions to be isomorphic. This can be stated in terms of the rank function δ_n that was introduced in [Tent 2000] generalizing the rank function for projective planes introduced in [Hall 1943]. It was used again in [Tent 2011; Müller and Tent 2019].

Definition 3.1. (i) For a finite graph $\Gamma = (V, E)$ with vertex set V and edge set E, define $\delta_n(\Gamma) = (n-1) \cdot |V| - (n-2) \cdot |E|$.

(ii) A (possibly infinite) graph Γ_0 is n-strong in some graph Γ, written $\Gamma_0 \leq_n \Gamma$, if and only if for all finite subgraphs X of Γ we have

$$\delta_n(X/X \cap \Gamma_0) := \delta_n(X) - \delta_n(X \cap \Gamma_0) \geq 0.$$

Remark 3.2. Note that δ_n is submodular, i.e., if $A \leq_n B$ and $C \subseteq B$, then $A \cap C \leq_n C$. Let A and B be finite graphs and let $E(A, B)$ denote the edges between elements of A and elements of B. Then

$$\delta_n(A/B) = \delta_n(A \setminus B) - (n-2)|E(A, B)|.$$

Remark 3.3 (cf. [Tent 2000, Lemma 2.4]). Let B be a graph which arises from the graph A by successively adding clean arcs between elements of distance $n+1$. Then $A \leq_n B$, $\delta_n(A) = \delta_n(B)$ and hence if $A \subseteq B \subseteq \Delta$ for some graph Δ with $A \leq_n \Delta$, then $B \leq_n \Delta$. In particular, if Γ_0 is a finite partial n-gon and $\Gamma = F(\Gamma_0) = \bigcup \Gamma_i$ as in Definition 2.4, then $\delta_n(\Gamma_i) = \delta_n(\Gamma_0)$ for all $i < \omega$. Hence any finite subset A_0 of Γ is contained in a finite subset $A \subseteq \Gamma$ with $\delta_n(A) = \delta_n(\Gamma_0)$.

Lemma 3.4 [Tent 2011, Proposition 2.5]. *Let Γ be a generalized n-gon which is generated by the graph Γ_0. The following are equivalent*:

(i) $\Gamma_0 \leq_n \Gamma$.

(ii) $\Gamma = F(\Gamma_0)$.

Remark 3.5. Note that for $k \geq n+3$ any finite subset A_0 of Γ^k is contained in a finite subset $A \subset \Gamma^k$ such that $\delta_n(A) = n-1+k = \delta_n(\gamma_k)$ and that $n-1+k$ is minimal with that property. Hence, if A and B are finite partial n-gons such that $\Gamma(A) \cong \Gamma(B)$, then $\delta_n(A) = \delta_n(B)$. In particular, if $\Gamma^k \cong \Gamma^m$, then $k = m$.

Definition 3.6. If $A \leq_n B$ are finite graphs such that $\delta_n(B/A) = 0$ and there is no proper subgraph $A \subset B' \subset B$ with $A \leq_n B' \leq_n B$ then B is called a *minimal* 0-extension.

Remark 3.7. Recall that $\mathcal{K}$ is the class of finite connected open partial n-gons. If $B \in \mathcal{K}$ is a minimal 0-extension of A, then either B is an extension of A by a clean arc of length $n-2$ or B is closed over A in the sense of Definition 2.16.

Lemma 3.8. *Let M be a model of T_n and A a finite subset of M. If $A \subseteq B \subseteq M$ and $\delta_n(B/A) \leq 0$, then B is algebraic over A.*

Proof. If $\delta_n(B/A) < 0$, then B is not HF-constructible over A and hence algebraic over A by Lemma 2.24. Now suppose that $\delta_n(B/A) = 0$. By submodularity we can decompose the extension B over A into a finite series $A = B_0 \leq_n B_1 \leq_n \cdots \leq_n B_k = B$, where each B_i is a minimal 0-extension of B_{i-1}. Hence it suffices to prove the claim for minimal 0-extensions and for such extensions the claim follows from Remark 3.7 and Lemma 2.24. □

The previous lemma directly implies:

Corollary 3.9. *Let Γ be an open generalized n-gon. If $A \subset \Gamma$ is such that every finite set $B_0 \supset A$ is contained in a finite set B such that $\delta_n(B) = \delta_n(A)$, then $\Gamma \subseteq \mathrm{acl}(A)$. In particular, any elementary embedding of Γ^k, $k \geq n+3$, into itself is surjective.*

Corollary 3.10 (cf. [Hyttinen and Paolini 2021, Corollary 6.3]). *For $n+3 \leq k, m \leq \omega$ we have $\Gamma^k \preccurlyeq \Gamma^m$ if and only if $k \leq m$.*

Proof. The direction from right to left is contained in Corollary 2.29. For the direction from left to right suppose Γ^k embeds elementarily into Γ^m for $m < k$ via f, so $f(\Gamma^m) \preccurlyeq f(\Gamma^k) \preccurlyeq \Gamma^m$. By Corollary 3.9 and the direction from right to left, we have $f(\Gamma^m) = \Gamma^m$, contradicting the fact that $\Gamma^m \subsetneq \Gamma^k$. □

To see that T_n has no prime model, we use results from [Tent 2000]. Hence we recall the definition of the class $\mathfrak{K}$ considered in [Tent 2000]. We show below that $\mathcal{K} \subseteq \mathfrak{K}$ and hence the results from [Tent 2000] apply.

Definition 3.11. Let $\mathfrak{K}$ be the class of finite partial n-gons A such that if A contains a $2k$-cycle for some $k > n$, then $\delta_n(A) \geq 2n+2$.

The following was shown in [Tent 2000, Lemma 3.12] (unfortunately the statement there contains a typo):

Lemma 3.12. *Let $A \in \mathfrak{K}$ with $|A| \geq n+2$. Then $\delta_n(A) \geq 2n$. Moreover, we have in fact $\delta_n(A) \geq 2n+2$, unless $|A| = n+2$ or A is an ordinary n-gon with either a path with $n-1$ new elements or a loose end attached.*

Proposition 3.13. *If $A \in \mathcal{K}$ contains a $2k$-cycle for some $k > n$, then $\delta_n(A) \geq 2n+2$. Hence $\mathcal{K} \subseteq \mathfrak{K}$.*

Proof. Let A be a minimal counterexample, so A contains a $2k$-cycle for some $k > n$ and $\delta_n(A) < 2n+2$. By minimality, A cannot contain a loose end, so $A = A_0 \cup \gamma$ for some clean arc γ. Then $\delta_n(A) = \delta_n(A_0) < 2n+2$. By minimality A_0 does not contain any $2k$-cycle for $k > n$ and hence $A_0 \in \mathfrak{K}$. By Lemma 3.12 we know that $|A_0| = n+2$ or A_0 is an ordinary n-gon with either a path with $n-1$ new elements or a loose end attached. But then $A = A_0 \cup \gamma$ does not contain any $2k$-cycle for $k > n$, a contradiction. □

Corollary 3.14. *If Γ is a generalized n-gon such that every finite set A_0 is contained in a finite set A with $\delta(A) = 2n+2$, then Γ does not contain any proper elementary submodels. In particular, Γ^{n+3} is minimal.*

Proof. This follows directly from Corollary 3.9 and Proposition 3.13. □

Definition 3.15. Consider Γ^{n+3} and choose copies $(\Gamma_i : i < \omega)$ of Γ^{n+3} such that $\Gamma_i \subsetneq \Gamma_{i+1}$. Put $\Gamma' = \bigcup_{i<\omega} \Gamma_i$.

Note that $\Gamma' \models T_n$ since T_n is an $\forall\exists$-theory. Also, every finite subset A_0 of Γ' is contained in a finite set $A \subset \Gamma'$ with $\delta_n(A) = 2n + 2$.

Proposition 3.16. *There exist open generalized n-gons which are not free. Specifically, $\Gamma' = \bigcup \Gamma_i$ is not free.*

Proof. Clearly Γ' is not finitely generated as any finite subset is contained in some Γ_i. So suppose towards a contradiction that Γ' is the free completion of an infinite hat-rack. Then for any $k \geq 2n + 2$ there exists a subset X of Γ' with $\delta_n(X) \geq k$ and $X \leq_n \Gamma'$, a contradiction to the observation that every finite subset A_0 of Γ' is contained in a finite set $A \subset \Gamma'$ with $\delta_n(A) = 2n + 2$. Thus Γ' is open and not free. □

Corollary 3.17. *The theory T_n of open generalized n-gons does not have a prime model.*

Proof. By Corollary 3.14, Γ^{n+3} and Γ' (as in Definition 3.15) have no proper elementary substructures. Since they are not isomorphic, this proves the claim. □

Since we can easily find (nonelementary) embeddings of Γ^m into Γ^k for $m \geq k$ we also obtain:

Corollary 3.18. *The theory of open generalized n-gons is not model complete and hence does not have quantifier elimination.*

Remark 3.19. Free ∞-gons are trees. Therefore the theory of free ∞-gons is in fact ω-stable as are their higher dimensional generalizations, right-angled buildings and free pseudospaces; see [Tent 2014].

Theorem 3.20 [Ammer 2022, Theorems 12.1 and 12.7]. *The theory T_n has weak elimination of imaginaries and is* 1*-ample, but not* 2*-ample.*

Furthermore, [Ammer 2022, Chapter 10] extends the proof from [Hyttinen and Paolini 2021] to obtain $2^{\aleph_0}$ many nonisomorphic countable open generalized n-gons for each n. Since T_n is not superstable, there are 2^κ many models of size κ for any uncountable κ.

4. Open vs. free

While we already saw in Proposition 3.16, that there are open generalized n-gons which are not free, we show in this final section that for finitely generated generalized n-gons the notions of open and free coincide. For $n = 3$ this was proved in [Hall 1943, Theorem 4.8].

Proposition 4.1. *Every finitely generated open generalized n-gon is free.*

For the proof we introduce the following concept:

Definition 4.2. We call partial n-gons A, B free-equivalent if $F(A) \cong F(B)$.

Lemma 4.3. *Let $\Gamma = F(A)$ be a generalized n-gon and suppose A is constructed from A_0 by first attaching a clean arc $\gamma = (x_1, \ldots, x_{n-2})$ and then attaching loose ends $z_1, \ldots, z_k$ whose respective (unique) neighbours belong to γ. Then there exist $z'_1, \ldots, z'_k \in \Gamma \setminus A$ with unique neighbours in A_0 such that A is free-equivalent to $A_0 \cup \{z'_1, \ldots, z'_k\}$.*

Proof. Let $\gamma' \subset A$ be a $2n$-cycle containing γ. Note that the opposites x'_i of x_i, $i = 1, \ldots, n-2$, in γ belong to A_0. By Remark 2.6 we can replace $z_i \in D_1(x_j)$ by the appropriate neighbour z'_i of the opposite x'_j of x_j and remove γ. □

Lemma 4.4. *Let $\Gamma = F(A)$ be a generalized n-gon and suppose A does not contain any cycle. Then there is a hat-rack B free-equivalent to A.*

Proof. Let $\gamma = (x_0, \ldots, x_k) \subset A$ be a simple path (i.e., without repetition of vertices) such that $k \geq n+3$ is maximal. The proof is by induction on the number of vertices of A not incident with γ. If A is a hat-rack, there is nothing to show. So let $a \in A$ have maximal distance from γ. If there is some $x_i \in \gamma$ such that $d(a, x_i) = n+1$, then let $a' \in \Gamma \setminus A$ be the unique neighbour of x_i with $d(a', a) = n-2$. Let A' be the graph obtained from A by replacing a by a'. Then $F(A') = F(A)$.

If there is no such vertex in γ, let $\gamma' \subset \Gamma$ be the clean arc connecting x_0 and x_{n+1}, so $F(A) = F(A \cup \gamma')$. There is some $y \in \gamma$ with $d(y, a) = n+1$. Let y' be the neighbour of y with $d(y', a) = n-2$. Then we replace A by $A' = (A \setminus \{a\}) \cup \gamma' \cup \{y'\}$. Thus $F(A) = F(A')$ and the claim follows from Lemma 4.3 and induction. □

We can now give the proof of Proposition 4.1:

Proof. Let Γ be an open generalized n-gon finitely generated over the finite partial n-gon A. We may assume that A is connected. If $\delta_n(A) = k$, then every finite set $A_0 \subset \Gamma$ is contained in a finite set A with $\delta(A) \leq k$. Hence we may assume that $A \leq_n \Gamma$ and so $\Gamma \cong F(A)$ by Lemma 3.4. Therefore it suffices to show that there is a finite hat-rack B free-equivalent to A.

By Lemma 2.18 consider a construction of A over the empty set. Clearly we may assume that the last construction step is the addition of a loose end. We now do induction over the number of steps adding a clean arc. If this number is zero, then A contains no cycles and the claim follows from Lemma 4.4. Now suppose A is obtained from A_0 by adding a clean arc γ and then adding a number of loose ends $z_1, \ldots, z_k$ (where the loose ends may be attached consecutively at a previous loose end). If all loose ends are incident with γ, then we finish using Lemma 4.3. Otherwise, we inductively reduce the distance of the loose ends by replacing them by a loose end at smaller distance to A: if z_i is a loose end, there is some $x \in A$ with $d(z_i, x) = n+1$ and such that x is not a loose end in A. Now replace z_i by the unique $z'_i \in D_1(x)$ with $d(z_i, z'_i) = n-2$. In this way we reduce to the case in Lemma 4.3 and finish. □

Remark 4.5. Using similar arguments one can also show that for finitely generated $\Gamma, \Gamma' \models T_n$ we have $\Gamma \cong \Gamma'$ if and only if $\Gamma = F(A)$, $\Gamma' = F(B)$ for finite A, B such that $\delta_n(A) = \delta_n(B)$. We leave the proof as an exercise for the interested reader.

References

[Ammer 2022] A.-M. Ammer, *Model theory of free and open generalized polygons*, Ph.D. thesis, Universität Münster, 2022.

[Feit and Higman 1964] W. Feit and G. Higman, "The nonexistence of certain generalized polygons", *J. Algebra* **1** (1964), 114–131. MR Zbl

[Hall 1943] M. Hall, "Projective planes", *Trans. Amer. Math. Soc.* **54** (1943), 229–277. MR Zbl

[Hyttinen and Paolini 2021] T. Hyttinen and G. Paolini, "First-order model theory of free projective planes", *Ann. Pure Appl. Logic* **172**:2 (2021), art. id. 102888. MR Zbl

[Kopeikina 1945] L. Kopeikina, "Free decompositions of projective planes", *Izv. Akad. SSSR. Ser. Mat.* **9** (1945), 495–526. In Russian. MR Zbl

[van Maldeghem 1998] H. van Maldeghem, *Generalized polygons*, Monographs Math. **93**, Birkhäuser, Basel, 1998. MR Zbl

[Müller and Tent 2019] I. Müller and K. Tent, "Building-like geometries of finite Morley rank", *J. Eur. Math. Soc.* (*JEMS*) **21**:12 (2019), 3739–3757. MR Zbl

[Siebenmann 1965] L. C. Siebenmann, "A characterization of free projective planes", *Pacific J. Math.* **15** (1965), 293–298. MR Zbl

[Tent 2000] K. Tent, "Very homogeneous generalized n-gons of finite Morley rank", *J. London Math. Soc.* (2) **62**:1 (2000), 1–15. MR Zbl

[Tent 2011] K. Tent, "Free polygons, twin trees, and CAT(1)-spaces", *Pure Appl. Math. Q.* **7**:3 (2011), 1037–1052. MR Zbl

[Tent 2014] K. Tent, "The free pseudospace is N-ample, but not $(N+1)$-ample", *J. Symb. Log.* **79**:2 (2014), 410–428. MR Zbl

[Tent and Ziegler 2012] K. Tent and M. Ziegler, *A course in model theory*, Lect. Notes Log. **40**, Assoc. Symbol. Log., La Jolla, CA, 2012. MR Zbl

[Tent and Zilber 2015] K. Tent and B. Zilber, "A non-desarguesian projective plane", preprint, 2015. arXiv 1510.06176

[Tits 1977] J. Tits, "Endliche Spiegelungsgruppen, die als Weylgruppen auftreten", *Invent. Math.* **43**:3 (1977), 283–295. MR Zbl

Received 6 Mar 2023. Revised 22 Sep 2023.

ANNA-MARIA AMMER:

am.ammer@web.de
Münster, Germany

KATRIN TENT:

tent@wwu.de
Universität Münster, Münster, Germany

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.449

New simple theories from hypergraph sequences

Maryanthe Malliaris and Saharon Shelah

Dedicated to Boris Zilber on the occasion of his 75th birthday.

We develop a family of simple rank one theories built over quite arbitrary sequences of finite hypergraphs. (This extends an idea from the recent proof that Keisler's order has continuum many classes, however, the construction does not require familiarity with the earlier proof.) We prove a model-completion and quantifier-elimination result for theories in this family and develop a combinatorial property which they share. We invoke regular ultrafilters to show the strength of this property, showing that any flexible ultrafilter which is good for the random graph is able to saturate such theories.

It is our pleasure to dedicate this to Boris for all the wonderful discoveries in model theory and its interaction with the rest of mainstream mathematics.

Recently, we proved that Keisler's order has continuum many pairwise incomparable classes, within the simple rank one theories [7]. A surprising point of that proof is that the theories built to obtain the continuum many incomparable classes can be very well understood, and are close to the random graph in various precise ways. So we can analyze carefully how their types are realized and omitted; this understanding helps in proving incomparability. Briefly, those theories were built over template sequences of growing finite graphs, and aspects of the combinatorics of the template graphs such as edge densities played a role in the behavior of types in the associated theories. This was a very nice interaction of the finite and the infinite, where the role of graphs seemed central; we should ask whether this understanding applies to a larger, significant family of simple theories. See also [2] and [9, Chapter VI] for context on Keisler's order.

In the present paper, we indeed find a way to extend ideas from the construction of the theories in [7] to build a nontrivial family of theories close to the random graph. Informally, the previous idea of using templates of sequences of growing finite graphs can be extended to templates of sequences of growing finite hypergraphs of

Research partially supported by NSF 1553653, and by an NSF-BSF award (NSF 2051825, BSF 3013005232). Paper 1206 in Shelah's list. We thank the referee for thoughtful comments on the manuscript.
MSC2020: 03C45.
Keywords: simple theories, model theory, Keisler's order.

any arity. We also indicate modifications of the construction involving equivalence relations rather than trees. Although we have found these theories in the context of investigating Keisler's order, indications are that they may be of general interest. Hence we have taken care to present them in a hopefully easily accessible way.

Meanwhile, an interesting aspect of Keisler's order on simple unstable theories is that it seems to be pointing the way towards isolating and analyzing an interesting family of theories "near" the random graph, which includes the incomparable theories of [7], and now the more general family developed here. We do not yet have indications whether this is *the* family. We do intend to look at whether the incomparability via ultraproducts can be carried out at the generality of these theories, and to consider other related questions in a future manuscript.

1. Templates and theories

To define our theories we first need to define a template, which is a growing sequence of finite hypergraphs, all of the same fixed arity k, satisfying certain mild conditions on the number of nodes and of edges. Our main case is $k > 2$, but the construction also makes sense for $k = 2$ (graphs) and so generalizes a slight variant[1] of the construction from [7]. The construction a priori makes sense without the conditions in Definition 1.4, but the model completion and quantifier elimination arguments use them. Given any such template, we then build a theory in a natural way.

Definition 1.1. Given a hypergraph (H, E), where E is a relation of arity k, say that k is the *arity* of the hypergraph.

Definition 1.2. Call a hypergraph (H, E) of arity k a *k-full hypergraph* if we can partition $E = E^* \cup E^{<k}$ such that (H, E^*) is a k-uniform hypergraph, meaning the edge relation is symmetric and irreflexive and holds only on tuples of k distinct elements, and $E^{<k}$ holds on *all* tuples with $< k$ distinct elements.

Informally, k-full hypergraphs are those obtained by starting with a k-uniform hypergraph, where the edge is symmetric and irreflexive and holds only on tuples of k distinct elements, and then extending it by setting the edge relation to hold on *all* tuples with repetition. (This is a technical help since nonedges in template hypergraphs indicate inconsistency in the related theory.) Note that it still is well defined to call k the arity of the hypergraph.

Definition 1.3. Given a hypergraph (H, E) of arity k, a *k-full-clique* is a set[2] $A \subseteq H$ where every sequence of k elements of A belongs to E, and a *k-independent set* is

[1] The reader familiar with the earlier paper will remember that the theories there were built on bipartite graphs, which had certain advantages for the ultrapower analysis. In order to extend to hypergraphs, rather than solving the problem of extending the bipartition to a multipartition, the problem was solved in a more satisfying way by eliminating the bipartition; then the extension to higher arities is even more natural.

a set $A \subseteq H$ with $\geq k$ members such that no sequence of k distinct elements of A belongs to E.

Definition 1.4. A *template* of arity k, $2 \leq k < \omega$, consists of a sequence $\mathcal{H} = \bar{\boldsymbol{h}} = \langle \boldsymbol{h}_n : n < \omega \rangle$ and a function $f_{\mathcal{H}} : \omega \to \omega \setminus \{0\}$ such that:

(0) $\lim_{n\to\infty} f_{\mathcal{H}}(n) = \infty$, meaning that for every $N < \omega$ there is $n < \omega$ such that $m \geq n \implies f_{\mathcal{H}}(m) \geq N$.

(1) for all $n < \omega$, $\boldsymbol{h}_n = (H_n, E_n)$ is a finite k-full hypergraph, $H_n = \|\boldsymbol{h}_n\|$ is a finite cardinal and so we identify the set of vertices H_n with the set $\{0, \ldots, H_n - 1\}$.

Moreover, for all $n < \omega$:

(2) $f_{\mathcal{H}}(n) \leq H_n < \aleph_0$.

(3) *(extension)* Let $t = f_{\mathcal{H}}(n)$. For every $i^0_0, \ldots, i^0_{k-2}, \ldots, i^{t-1}_0, \ldots, i^{t-1}_{k-2}$ from H_n, there exists $s \in H_n$ such that $\langle s, i^\ell_0, \ldots, i^\ell_{k-2} \rangle \in E_n$ for all $\ell < t$.

We say $\mathcal{H}$ is a template if $(\mathcal{H}, f)$ is for some f.

Remark 1.5. For notational simplicity in Definition 1.4, we fix k. We could also have defined a parameter $\boldsymbol{k}_n$ for each n measuring the fullness.

Definition 1.6. A *template* is a template of arity k for some $k < \omega$.

For example, the sequence of hypergraphs given by $H_n = n + 1$ and $E_n = {}^k H_n$ is a template of arity k. For a more interesting example, choose the $\boldsymbol{h}_n$ to be a sequence of finite random hypergraphs, with size and edge probability sufficient to give the extension condition Definition 1.4(3). For a similar sufficient calculation in the original case of graphs, see [7, §6].

As the next definition suggests, it will be useful to think of trees naturally associated to paths through the template hypergraphs.

Definition 1.7. Given a template $\mathcal{H}$, and recalling H_n from Definition 1.4, define

$$X_{\mathcal{H}} := \{\rho : \rho \in {}^{\omega>}\omega, 0 \leq \rho(n) < H_n \text{ for all } n < \operatorname{lgn}(\rho)\}$$

to be, informally, the set of finite sequences of choices of vertices from initial segments of our hypergraph sequence, naturally forming a tree. Define

$$\operatorname{leaves}(X_{\mathcal{H}}) = \{\rho \in {}^{\omega}\omega : \eta_i \restriction n \in X_{\mathcal{H}} \text{ for all } n < \omega\}$$

to be the "limit points" of this set.

Definition 1.8. We define a theory $T_0 = T_0(\mathcal{H})$ based on the template $\mathcal{H}$ to be the following universal theory in the following language.

---

[2]In the interesting case, a set with $\geq k$ members, but this hypothesis is not strictly needed as the sequences can contain repetitions. In the case of the independent set, we need $|A| \geq k$ and could have asked $|A| > k$.

(1) $\mathcal{L} = \mathcal{L}_{\mathcal{H}}$ contains equality, a k-place relation R, and countably many unary predicates

$$\{Q_\eta : \eta \in X_{\mathcal{H}}\}.$$

(2) T_0 contains universal axioms stating that R is a symmetric k-uniform hypergraph, i.e., R holds only on distinct k-tuples and if it holds on some k-tuple it holds on all its permutations.[3]

(3) If $\eta \trianglelefteq \nu \in X_{\mathcal{H}}$ then T_0 contains the axiom

$$(\forall x)(Q_{\langle\rangle}(x)) \wedge (\forall x)(Q_\nu(x) \implies Q_\eta(x))$$

saying that $Q_{\langle\rangle}$ names everything, and Q_ν refines Q_η.

(4) If $\eta \in X_{\mathcal{H}}$, $\operatorname{lgn}(\eta) = m$ and $i \neq j < \|\boldsymbol{h}_m\|$ then T_0 contains the axiom

$$(\forall x)\big(\neg(Q_{\eta^\frown\langle i\rangle}(x) \wedge Q_{\eta^\frown\langle j\rangle}(x))\big).$$

Moreover, T_0 contains the axiom $(\forall x)\big(Q_\eta(x) \implies \bigvee_i Q_{\eta^\frown\langle i\rangle}(x)\big)$, so the predicates $\langle Q_{\eta^\frown\langle i\rangle} : i < \|\boldsymbol{h}_m\|\rangle$ partition Q_η.

(5) For every $\eta_0, \dots, \eta_{k-1}$ from $X_{\mathcal{H}}$ and $n < \min\{\operatorname{lgn}(\eta_0), \dots, \operatorname{lgn}(\eta_{k-1})\}$, if $\langle \eta_0(n), \dots, \eta_{k-1}(n)\rangle \notin E_n$ then T_0 contains the axiom

$$(\forall x_0, \dots, x_{k-1})(P_{\eta_0}(x_0) \wedge \cdots \wedge P_{\eta_{k-1}}(x_{k-1}) \implies \neg R(x_0, \dots, x_{k-1}))$$

forbidding any edges across these predicates.

Discussion 1.9. Informally, the unary predicates give a model $M \models T_0$ the (hard-coded) structure of a tree. We have $\forall x\, Q_{\langle\rangle}(x)$. The model is first partitioned into predicates $Q_{\langle i\rangle}$ for $i < \|\boldsymbol{h}_0\|$. By induction on $m \geq 1$, each predicate Q_η (where $\operatorname{lgn}(\eta) = m$, i.e., η is a function with domain $\{0, \dots, m-1\}$) is partitioned into $\|\boldsymbol{h}_m\|$ disjoint pieces, the $Q_{\eta^\frown\langle i\rangle}$. So any $a \in M$ will be in some concentric sequence of predicates $\langle Q_{\rho\restriction n} : \rho \in \operatorname{leaves}(X_{\mathcal{H}}), n < \omega\rangle$. Call ρ the *leaf* of a (see Definition 1.10). Note that we have arranged our indexing so that, in this notation, if $\rho(n) = i$ we have

$$a \in Q_{(\rho\restriction n)^\frown\langle i\rangle},$$

in other words, that its predicate at level n corresponds to the i-th element of H_n. The final condition on edges amounts to the following. Given $a_0, \dots, a_{k-1}$ in a model $M \models T_0$, each element a_i belongs to some leaf ρ_i, and an edge R cannot occur on $\langle a_0, \dots, a_{k-1}\rangle$ *unless* for every $n < \omega$, $\langle \rho_i(n) : i < k\rangle$ is an edge in E_n. (Since T_0 is a universal theory, of course, it records here just what is forbidden, and remains agnostic about whether edges do occur if permitted; a model completion, such as we shall construct soon, would have more information.)

[3] Note that R, the hypergraph relation in the theory, is symmetric irreflexive, while E_n, the hypergraph relations in the templates, need not be irreflexive by the definition of "k-full".

Notice the "sparsification" of edges, or rather the accumulation of rules forbidding edges, as we go deeper into the "tree". If $\eta_0, \dots, \eta_{k-1}$ are elements of $X_{\mathcal{H}}$ of length $m+1$, and $\langle \eta_0(m), \dots, \eta_{k-1}(m) \rangle \notin E_m$, then in M we know there can be no R-edges spanning elements chosen from the predicates $Q_{\eta_0}, \dots, Q_{\eta_{k-1}}$ regardless of how these elements sit in subsequent predicates. If on the other hand $\langle \eta_0(\ell), \dots, \eta_{k-1}(\ell) \rangle \in E_\ell$ for $\ell \leq m$, then a priori there may be edges spanning *some* elements from the predicates $Q_{\eta_0}, \dots, Q_{\eta_{k-1}}$, but it may depend a priori on how those elements sit in subsequent predicates and what the templates say there.

The following auxiliary objects may clarify the picture.

Definition 1.10. Fix a template $\mathcal{H}$ of arity k. Let $T_0 = T_0(\mathcal{H})$ and let $M \models T_0$.

(1) For $a \in M$, define $\operatorname{leaf}(a)$ to be the unique $\rho \in \operatorname{leaves}(X_{\mathcal{H}})$ such that

$$M \models a \in Q_{\rho \restriction n} \quad \text{for all } n < \omega.$$

(2) Let $\boldsymbol{h}_\infty$ be the k-uniform hypergraph with vertex set $H_\infty := \operatorname{leaves}(X_{\mathcal{H}})$ and with edge relation E_∞ given by

$$\langle \rho_0, \dots, \rho_{k-1} \rangle \in E_\infty \iff \langle \rho_0(n), \dots, \rho_{k-1}(n) \rangle \in E_n \text{ for all } n < \omega.$$

Of course $\boldsymbol{h}_\infty = \boldsymbol{h}_\infty(\mathcal{H})$.

Observation 1.11. Definition 1.8(5) implies that if $M \models T_0$, $a_0, \dots, a_{k-1} \in M$, we can have $M \models R(a_0, \dots, a_{k-1})$ only if $\langle \operatorname{leaf}(a_0), \dots, \operatorname{leaf}(a_{k-1}) \rangle \in E_\infty$.

Example 1.12. Suppose that $k = 3$, $\langle 0, 1, 2 \rangle \in E_0$ and $\langle 3, 4, 5 \rangle \in E_1$. Then R-edges are not a priori forbidden in T_0 between $Q_{\langle 03 \rangle}, Q_{\langle 14 \rangle}, Q_{\langle 25 \rangle}$, nor between $Q_{\langle 04 \rangle}, Q_{\langle 15 \rangle}, Q_{\langle 23 \rangle}$ remembering symmetry of E_1, nor between $Q_{\langle 01 \rangle}, Q_{\langle 10 \rangle}, Q_{\langle 20 \rangle}$ remembering E_1 is k-full.

2. Model completion and quantifier elimination

Convention 2.1. For the entirety of this section, fix a template $\mathcal{H}$, $f_{\mathcal{H}}$ of arity $k \geq 2$, and thus $\boldsymbol{h}_\infty$ and T_0 as in Definitions 1.10 and 1.8, respectively.

Claim 2.2. *For any $\rho \in H_\infty$ there are continuum many tuples $\rho_0, \dots, \rho_{k-2} \in H_\infty$ such that $\langle \rho, \rho_0, \dots, \rho_{k-2} \rangle \in E_\infty$, i.e., each leaf in this graph is contained in continuum many edges.*

Proof. By extension (Definition 1.4(3)). $\square$

The next claim is a key use of Definition 1.4(3): in some sense, it shows that consistency in the template at large enough finite levels can be extended to full consistency.

Claim 2.3 (completion to a type). *For any $1 \leq t < \omega$ and any choice of t k-tuples $\rho^0_0, \ldots, \rho^0_{k-2}, \ldots, \rho^{t-1}_0, \ldots, \rho^{t-1}_{k-2}$ from H_∞, **if** there exists $\nu \in X_{\mathcal{H}}$ such that* $\operatorname{lgn}(\nu) > \min\{m : f(n) \geq t$ *for all* $n \geq m\}$ *and*

$$\langle \nu(\ell), \rho^i_0(\ell), \ldots, \rho^i_{k-2}(\ell)\rangle \in E_\ell \quad \textit{for all } i < m \textit{ and } \ell < \operatorname{lgn}(\nu),$$

***then** we can choose ν_* such that $\nu \trianglelefteq \nu_* \in H_\infty$ and*

$$\langle \nu_*(\ell), \rho^i_0(\ell), \ldots, \rho^i_{k-2}(\ell)\rangle \in E_\infty \quad \textit{for all } i < m \textit{ and } \ell < \omega.$$

Proof. Let $n := \operatorname{lgn}(\nu)$. By induction on $r < \omega$ let us prove that we can find $\nu_r \in X_{\mathcal{H}}$ of length $n + r$ such that $\nu \trianglelefteq \nu_r$ and

$$\langle \nu_r(\ell), \rho^i_0(\ell), \ldots, \rho^i_{k-2}(\ell)\rangle \in E_\ell \quad \text{for all } i < m \text{ and } \ell < n + r.$$

For $\ell = 0$ take $\nu_t = \nu$. For $\ell > 0$, apply extension (Definition 1.4(3)) to the tuples

$$\rho^0_0(n+r-1), \ldots, \rho^0_{k-2}(n+r-1), \ldots, \rho^{t-1}_0(n+r-1), \ldots, \rho^{t-1}_{k-2}(n+r-1)$$

in the hypergraph $\boldsymbol{h}_{n+r-1}$ and let b be the appropriate element of H_{n+r-1} returned by that axiom. Then $\nu_r := \nu_{r-1}{}^\frown\langle a\rangle$ fits the bill. $\square$

Definition 2.4. For any $m < \omega$, define T^m_0 to be the restriction of T_0 to the language with equality, a k-place relation R, and unary predicates

$$\{Q_\eta : \eta \in X_{\mathcal{H}}, \operatorname{lgn}(\eta) \leq m\}.$$

Claim 2.5. *For each $m < \omega$, the model completion T^m of T^m_0 exists.*

Proof. Just as in the case of graphs [7, Observation 2.16], each T^m_0 is a universal theory in a finite relational language. The class of its models has the joint embedding property JEP for any two M_1, M_2 with $|M_1| \cap |M_2| = \varnothing$, and the amalgamation property AP when we have models M_1, M_2 and M_0 with $M_0 \models T^m_0$ and $M_0 \subseteq M_\ell$ for $\ell = 1, 2$ and $|M_1| \cap |M_2| = |M_0|$. To see this in both cases, the model N whose domain is $|M_1| \cup |M_2|$, such that $Q^N = Q^{M_1} \cup Q^{M_2}$ for each unary predicate Q and $R^N = R^{M_1} \cup R^{M_2}$ for the edge relation R, will be a model of T^m_0. Thus T^m exists. $\square$

Remark 2.6. Regarding the model completion, if $M \models T^m$, then M is infinite, and indeed for each unary predicate $Q \in \tau(T^m)$, Q^M is infinite. Moreover,[4] for any $\eta_0, \ldots, \eta_{k-1} \in X_{\mathcal{H}}$ with $\operatorname{lgn}(\eta_\ell) = m$ for $\ell < k$:

(a) If $(\eta_0(i), \ldots, \eta_{k-1}(i)) \in E_i$ for all $i < m$, then R^M on $Q^M_{\eta_0} \times \cdots \times Q^M_{\eta_{k-1}}$ "is a random hypergraph" in the sense of first-order logic, meaning that *if* $A \subseteq Q^M_{\eta_1} \times \cdots \times Q^M_{\eta_{k-1}}$ and $B \subseteq |M|^k$ and[5] "$A \cap B = \varnothing$" in the strong sense

[4] We can extend case (a) to $\{\eta_0, \ldots, \eta_{\ell-1}\}$ for some larger finite ℓ which form a k-full-clique in the same strong hereditary sense.

[5] We could have asked that $B \subseteq Q^M_{\eta_1} \times \cdots \times Q^M_{\eta_{k-1}}$, but the stronger statement is true.

that no permutation of any $(a_1, \dots, a_{k-1}) \in A$ belongs to B, *then* the set of formulas

$$p(x) = \{R(x, a_1, \dots, a_{k-1}) : (a_1, \dots, a_{k-1}) \in A\} \\ \cup \{\neg R(a, b_1, \dots, b_{k-1}) : (b_1, \dots, b_{k-1}) \in B\}$$

is a partial type in M, so in particular is realized if A, B are both finite.[6]

(b) If not, then $(Q^M_{\eta_0} \times \cdots \times Q^M_{\eta_{k-1}}) \cap R^M = \varnothing$.

Next we upgrade [7, Claim 2.17] to the context of hypergraphs.

Notation 2.7. For an ordered set X, let $\mathrm{inc}_\ell(X)$ be the set of strictly increasing ℓ-tuples of elements of X.

Definition 2.8. Given T^m for some $m < \omega$ and $M, N \models T^m$, recall that

(1) $a_1, \dots, a_n \in |M|$ and $b_1, \dots, b_n \in |N|$ have the same quantifier-free $\tau(T^m)$-type when they agree on equality, instances of R, and predicates Q_η up to $\mathrm{lgn}(\eta) = m$.

(2) $\varphi(x, y_1, \dots, y_n)$ is a *complete* quantifier-free formula of $\tau(T^m)$ when
   (a) for every unary predicate $Q \in \tau(T^m)$ and variable $z \in \{x, y_1, \dots, y_n\}$, either $\varphi \vdash Q(z)$ or $\varphi \vdash \neg Q(z)$;
   (b) for every z_0, z_1 from $\{x, y_1, \dots, y_n\}$, either $\varphi \vdash z_0 = z_1$ or $\varphi \vdash z_0 \neq z_1$;
   (c) for every $z_0, \dots, z_{k-1}$ from $\{x, y_1, \dots, y_n\}$, either $\varphi \vdash R(z_0, \dots, z_{k-1})$ or $\varphi \vdash \neg R(z_0, \dots, z_{k-1})$.

   Recall that the language is finite so this is well defined.

Our next lemma says that for each m, the truth of sentences of $\tau(T^m_0)$ of length $\leq m$ soon stabilizes in the sequence of theories T^k as k goes to infinity.

Lemma 2.9. *For every $m < \omega$, the following holds. Let*

$$m_* \geq \min\{n : n' \geq n \Longrightarrow f_{\mathcal{H}}(n') \geq m\}.$$

If $M \models T^{m_*}$, $N \models T^{m_*+1}$ and φ is a sentence of $\tau(T^m)$ of length $\leq m$, then $M \models \varphi \Longleftrightarrow N \models \varphi$.

Proof. To prove the lemma by induction on complexity of formulas, it suffices to show the following:

[6] Note that by our assumption of the template hypergraphs being "k-full", we are in case (a) whenever $|\{\eta_0, \dots, \eta_{k-1}\}| < k$. The hypergraph edge R is a k-uniform hypergraph in M, of course, so any $(a_0, \dots, a_{k-1}) \in R^M$ will be a tuple of distinct elements, but fullness of the template hypergraphs means some of the elements in such a tuple are a priori allowed to come from the same predicate at any given level. In particular, for each $\eta \in X_{\mathcal{H}}$ with $\mathrm{lgn}(\eta) = m$, $(Q^M_\eta, R^M \restriction Q^M_\eta)$ is a random k-ary hypergraph in the usual sense of first-order logic.

Suppose $\varphi(x, y_1, \ldots, y_n)$ is a complete quantifier-free formula of $\tau(T^m)$ of length $\leq m$, so note $n < m$. Suppose $a_1, \ldots, a_n \in |M|$ and $b_1, \ldots, b_n \in |N|$ have the same quantifier-free $\tau(T^m)$-type. Then there exists $a \in |M|$ such that $M \models \varphi(a, a_1, \ldots, a_n)$ ***if and only if*** *there exists $b \in |N|$ such that $N \models \varphi(b, b_1, \ldots, b_n)$.* $(\oplus)$

Without loss of generality, the sequences $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are without repetition.

For left to right, suppose that $a \in |M|$ exists and $a \in \{a_1, \ldots, a_n\}$; otherwise it is trivial. We will need notation to record edges and nonedges made by a. For $\bar{\imath}$ any sequence of elements of $\{1, \ldots, n\}$, denote by $\bar{a}_{\bar{\imath}}$ the sequence $\langle a_{\bar{\imath}(\ell)} : \ell < \operatorname{lgn}(\bar{\imath})\rangle$. Let

$$C = \{\bar{\imath} = \langle i_0, \ldots, i_{k-2}\rangle : \bar{\imath} \in \operatorname{inc}_{k-1}(\{1, \ldots, n\}), \langle a\rangle^\frown \bar{a}_{\bar{\imath}} \in R^M\}$$

represent the set of R-edges made by a to $\{a_1, \ldots, a_{n-1}\}$. Note that $|C| < n^k$. Correspondingly, let

$$D = \operatorname{inc}_{k-1}(\{1, \ldots, n\}) \setminus C$$

represent the set of non-R-edges made by a to $\{a_1, \ldots, a_{n-1}\}$. If $C = \varnothing$ finding a corresponding b is immediate, so assume $C \neq \varnothing$.

Each element c of M belongs to a unique predicate Q_η with $\operatorname{lgn}(\eta) = m_*$; call it "the m_*-leaf of c" and write $\operatorname{leaf}_{m_*}(c) = \eta$. Let $\rho = \operatorname{leaf}_{m_*}(a)$ and let $\rho_i = \operatorname{leaf}_{m_*}(a_i)$ for $i = 1, \ldots, n$. The definition of $T_0^{m_*}$ and the existence of a tell us that necessarily

$$\text{for every } \iota = \langle i_0, \ldots, i_{k-2}\rangle \in C, \text{ for every } \ell < m_*,$$
$$\langle \rho(\ell), \rho_{i_0}(\ell), \ldots, \rho_{i_{k-2}}(\ell)\rangle \in E_\ell.$$

Meanwhile, each element d of N belongs to a unique predicate Q_η with $\operatorname{lgn}(\eta) = m_* + 1$; write $\operatorname{leaf}_{m_*+1}(d) = \eta$. So let $\nu_i = \operatorname{leaf}_{m_*+1}(b_i)$ for $i = 1, \ldots, n$. Note that $\operatorname{leaf}_{m_*}$ and $\operatorname{leaf}_{m_*+1}$ a priori depend on the models M and N, but by our assumption that $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ have the same quantifier-free $\tau(T_m)$-type, necessarily $\nu_i \upharpoonright m_* = \rho_i$ for $i = 1, \ldots, n$. Apply extension (Definition 1.4(3)) to the set of $(k-1)$-tuples

$$\{\langle \nu_{i_0}(\ell), \ldots, \nu_{i_{k-2}}(\ell)\rangle : \bar{\imath} = \langle i_0, \ldots, i_{k-2}\rangle \in C\},$$

recalling our choice of m_*, and let s be the element of H_{m_*} returned. Define $\nu = \eta^\frown\langle s\rangle$. Now we have that

$$\text{for every } \iota = \langle i_0, \ldots, i_{k-2}\rangle \in C, \text{ for every } \ell < m_* + 1,$$
$$\langle \nu(\ell), \nu_{i_0}(\ell), \ldots, \nu_{i_{k-2}}(\ell)\rangle \in E_\ell.$$

So by definition of $T_0^{m_*+1}$, $\varphi(x, b_1, \ldots, b_n)$ is consistent with N, and b exists because N is model complete.

The other direction, right to left, is simpler. Suppose that $b \in |N|$ exists and $b \notin \{b_1, \ldots, b_n\}$. As before, define C to be the set of representatives of edges. Suppose $\mathrm{leaf}_{m_*+1}(b) = \nu$ and $\mathrm{leaf}_{m_*+1}(b_i) = \nu_i$. Then since b exists and N is a model of $T_0^{m_*+1}$, necessarily

$$\text{for every } \iota = \langle i_0, \ldots, i_{k-2}\rangle \in C, \text{ for every } \ell < m_* + 1,$$
$$\langle \nu(\ell), \nu_{i_0}(\ell), \ldots, \nu_{i_{k-2}}(\ell)\rangle \in E_\ell.$$

A fortiori, then,

$$\text{for every } \iota = \langle i_0, \ldots, i_{k-2}\rangle \in C, \text{ for every } \ell < m_*,$$
$$\langle \nu(\ell), \nu_{i_0}(\ell), \ldots, \nu_{i_{k-2}}(\ell)\rangle \in E_\ell,$$

so by definition of $T_0^{m_*}$, $\varphi(x, a_1, \ldots, a_n)$ is consistent with M, and since it is complete $\varphi \vdash Q_{(\nu\restriction m_*)}(x)$, and a exists because M is model complete. ☐

Corollary 2.10. *"The limit theory of $\langle T^m : m < \omega\rangle$ is well defined and is a complete, model complete theory which extends T_0." For every $m < \omega$ and every formula φ of $\tau(T^m)$ in at least one free variable,*[7] *for some quantifier-free formula ψ of $\tau(T^m)$, for every n large enough, we have that*

$$(\forall \bar{x})(\varphi(\bar{x}) \equiv \psi(\bar{x})) \in T^n.$$

Lemma 2.11. *The theory T is simple rank* 1.

Proof. Assume for a contradiction that $\langle \bar{a}_i : i < \kappa\rangle$, $\kappa = \mathrm{cof}(\kappa) \geq (2^{\aleph_0})^+$ witnesses that some formula $\varphi(\bar{x}, \bar{y})$ n-divides, in a large κ-saturated model $M \models T$. Without loss of generality, possibly adding dummy variables, $\mathrm{lgn}(\bar{x}) = \mathrm{lgn}(\bar{y}) =: m$.

For each $i < \kappa$, let $\bar{b}_i$ be such that $M \models \varphi[\bar{b}_i, \bar{a}_i]$. Since κ is large enough (i.e., since $\mathrm{cof}(\kappa) > 2^{\aleph_0}$), for some $\mathcal{U} \in [\kappa]^\kappa$, for each $\ell < m$ there is $\nu_\ell \in H_\infty$ such that $\mathrm{leaf}(\bar{b}_{i,\ell})$ is constantly equal to ν_ℓ, and there is $\rho_\ell \in H_\infty$ such that $\mathrm{leaf}(\bar{a}_{i,\ell})$ is constantly equal to ρ_ℓ.

Let φ' be an extension of φ which is complete for $\{=, R\}$ (it will obviously only contain information about unary predicates up to some finite level) such that $M \models \varphi'[\bar{b}_i, \bar{a}_i]$ for $i \in \mathcal{V} \in [\mathcal{U}]^\kappa$. We may assume φ' is quantifier-free. Without loss of generality, φ' does not imply any instances of equality among the x's or between the x's and the y's. In what follows, replace φ by φ' and $\langle \bar{a}_i : i < \kappa\rangle$ by $\langle \bar{a}_i : i \in \mathcal{V}\rangle$.

We would like to show that

$$\Sigma(\bar{x}) = \{\varphi(\bar{x}, \bar{a}_i) : i < \kappa\} \text{ is consistent.}$$

It suffices by induction on $j < m$ to choose elements b_j so that b_j realizes the set of formulas $\Sigma^j(b_0, \ldots, b_{j-1}, x_j)$, where Σ^j is the restriction of Σ to the

[7] Since we do not have constants in the language.

variables $x_0, \dots, x_j$. In the case $\ell(\bar{x}) = 1$, write $\nu = \mathrm{leaf}(x)$, and this case follows from three simple observations:

- φ is without loss of generality quantifier-free; we assumed no instances of equality between the x's, and our theory has no algebraicity.
- The template hypergraphs contribute no restriction to this set of formulas, since if $R(x, a_{j_0}, \dots, a_{j_{k-2}})$ is implied by Σ then we know by our construction that $(\nu, \rho_{j_0}, \dots, \rho_{j_{k-2}}) \in E_\infty$.
- The indiscernibility of $\langle \bar{a}_i : i < \kappa \rangle$, transitivity of equality, and consistency of each instance $\varphi(x, \bar{a}_i)$ together mean that if $R(x, a_{j_0}, \dots, a_{j_{k-2}})$ is implied by Σ and $\neg R(x, a_{\ell_0}, \dots, a_{\ell_{k-2}})$ is implied by Σ, then no permutation of $\langle a_{j_0}, \dots, a_{j_{k-2}} \rangle$ is equal to $\langle a_{\ell_0}, \dots, a_{\ell_{k-2}} \rangle$ (so the "positive" and "negative" edges required by Σ cause no explicit contradiction).

Observe that the inductive step, since we will have already chosen the earlier values b_ℓ ($\ell < j$), will reduce to the case $\mathrm{lgn}(\bar{x}) = 1$ (using $\mathrm{lgn}(\bar{y}) = m + j$). This is enough to deduce the consistency of Σ, so there is no dividing. $\square$

Conclusion 2.12. *Given any template $\mathcal{H}$, the universal theory $T_0 = T_0(\mathcal{H})$ has a model completion $T = T(\mathcal{H})$ which is well defined, eliminates quantifiers, is simple rank* 1, *and is equal to the limit of $\langle T^m : m < \omega \rangle$.*

Discussion 2.13. We could have defined the theory to be "based on" predicates naming classes of crosscutting finite equivalence relations, rather than levels of trees, in the natural way. Alternatively, we could make E_n be a k-place relation on $\prod_{\ell \le n} H_\ell$.

3. A combinatorial property

In this section we give Definition 3.1, which is supposed to capture what is simple about the theories of Section 1, not necessarily what is complicated about them. In Section 4 we shall use this to give a sufficient condition for ultrafilters to saturate such theories. First let us motivate the property.

Suppose, with no assumptions on T or φ, we have a sequence of instances of φ

$$\varphi(\bar{x}, \bar{a}_0), \dots, \varphi(\bar{x}, \bar{a}_{s-1})$$

forming a partial type, and suppose we replace each $\bar{a}_i$ by a sequence $\bar{b}_i$ having the same type over the empty set. (We don't ask that $\bar{a}_i$ and $\bar{a}_j$ have the same type for $i \neq j$, just that $\bar{a}_i$ and $\bar{b}_i$ have the same type for each i.) Then a priori,

$$\varphi(\bar{x}, \bar{b}_0), \dots, \varphi(\bar{x}, \bar{b}_{s-1})$$

need not remain a partial type. An example is $\varphi(x; y_0, y_1) = y_0 < x < y_1$ in the theory of dense linear orders: any two pairs of increasing elements have the same type over the empty set, but we can choose the $\bar{a}$'s to be a sequence of intervals which are concentric, and the $\bar{b}$'s a sequence which are disjoint. Similar examples arise whenever we have a tuple beginning two indiscernible sequences, one which witnesses dividing of φ and one which does not.

An example of (T, φ) where such a substitution *does* remain a partial type, for trivial reasons, is the theory $T = T_{\mathbf{rg}}$ of the random graph, and $\varphi(x, y) = R(x, y)$, using only the positive instance. Note that $\varphi(x; y, z) = R(x, y) \wedge \neg R(x, z)$ would not work, however, since in changing from $\bar{a}$'s to $\bar{b}$'s we could introduce collisions among the parameters. A less trivial example is the positive instance of the edge relation in the theories of Section 1, which in fact satisfy a stronger condition, (as does the random graph), as we shall now see.

Among the examples of (T, φ) where this *does* work, we can ask just how much of each type we need to preserve when changing the parameters from $\bar{a}_i$'s to $\bar{b}_i$'s. Rather than preserving all formulas, perhaps it would be sufficient to enumerate some formulas of the type of each parameter in some coherent way, and then preserve some finite initial segment of each of these lists. It is reasonable that the length of the initial segment needed would depend on s, the number of instances we are dealing with. This is essentially what the next definition says.[8]

Definition 3.1. We say that $(T, \varphi(\bar{x}, \bar{y}))$ has the *pseudo-nfcp* when T is countable and we can assign to each type $p \in \mathcal{P}$, where

$$\mathcal{P} := \{p : p \in \boldsymbol{S}_{\ell(\bar{y})}(\varnothing) \text{ and } p \text{ contains the formula } \exists\bar{x}\, \varphi(\bar{x}, \bar{y})\},$$

a function $f_p : \omega \to \omega$ such that

(1) (continuity) for each $m < \omega$, if $f_p(m) = r$, then for some $\psi(\bar{y}) \in p$, for any other $q \in \mathcal{P}$, if $\psi \in q$, then $f_q(m) = r$.

(2) For notational convenience, if $p = \mathrm{tp}(\bar{a}) \in \mathcal{P}$, we may write $f_{\bar{a}}$ for f_p.

(3) For every $s \geq 1$ there is $n < \omega$ such that whenever $\bar{a}_0, \ldots, \bar{a}_{s-1}, \bar{b}_0, \ldots, \bar{b}_{s-1}$ are sequences from $\mathfrak{C}_T$, hence each realizing types in $\mathcal{P}$, and

$$f_{\bar{a}_\ell} \restriction n = f_{\bar{b}_\ell} \restriction n \quad \text{for all } \ell < s$$

and $\{\varphi(\bar{x}, \bar{a}_\ell) : \ell < s\}$ is a partial type, then $\{\varphi(\bar{x}, \bar{b}_\ell) : \ell < s\}$ is also a partial type. In the proofs that follow, we will refer to this by saying "(T, φ) is (s, n)-compact."

Discussion 3.2. (1) So Definition 3.1 is a kind of compactness demand, that is, given $(T, \varphi(\bar{x}, \bar{y}))$, to know if $\mathfrak{C}_T \models (\exists\bar{x}) \bigwedge_{\ell<s} \varphi(\bar{x}, \bar{b}_\ell)$ we need to know just finite approximations to the type of each $\bar{b}_\ell$ (not of $\bar{b}_0^\frown \cdots \bar{b}_{s-1}$!) and the size of "finite", represented here by n, depends just on s (and on T and φ).

[8]The provisional name is because it captures a key property of theories from [7].

(2) We could have defined the range of each function f to be finite subsets of ω, as would be convenient in Claim 3.4, or a more complicated set (of bounded, say countable, size); or we could have used $\{0, 1\}$.

(3) We could extend the definition to uncountable theories with more work.

Remark 3.3. In the context of Definition 3.1, when that definition is satisfied, we may define two functions F and G as follows.

(a) Define $F : \omega \to \omega$ by

$$s \mapsto \min\{n < \omega : (T, \varphi) \text{ is } (s, n)\text{-compact}\}$$

which expresses that in order for s instances to remain consistent, their functions f must be preserved at least up to $F(s)$. This is well defined since we assume the definition is satisfied. There are two cases:

(1) $\lim_{s\to\infty} F(s) \to \infty$.

(2) $\lim_{s\to\infty} F(s) = N < \infty$.

(b) Define $G : \omega \to \omega \cup \{\infty\}$ by $n \mapsto \infty$ if (T, φ) is (s, n)-compact for all $n < \omega$, and otherwise by $n \mapsto \max\{s < \omega : (T, \varphi) \text{ is } (s, n)\text{-compact}\}$, which expresses that if the functions f are preserved up to n then $G(n)$ instances can safely remain consistent. Here

$$\lim_{n\to\infty} G(n) = \infty, \tag{$\star$}$$

possibly attaining the limit already at some finite n.

Claim 3.4. *Let T be one of the theories from Section 1, built from $\mathcal{H}$, $f_{\mathcal{H}}$ of arity k. Let $\varphi(x, y_0, \ldots, y_{k-2}) = R(x; y_0, \ldots, y_{k-2})$. Then (T, φ) has the pseudo-nfcp.*

Proof. In this context, by quantifier elimination, the set of 1-types over the empty set are the set of "leaves", that is, each 1-type is specified by choosing some $\eta \in \text{leaves}(X_{\mathcal{H}})$ and considering $\{Q_{\eta\restriction n} : n < \omega\}$.

If $k = 2$, this also specifies $\mathcal{P}$. Otherwise, specifying a type $p(y_0, \ldots, y_{k-2}) \in \mathcal{P}$ involves specifying the leaf of each y_i, and if two elements share the same leaf, whether they are equal.

Consider any enumeration $\langle \psi_i : 1 \le i < \omega\rangle$ of the predicates $Q_\eta(y)$ of $\tau(T)$ which enumerates in nondecreasing order of $\text{lgn}(\bar{\eta})$. Fix also in advance an enumeration of the subsets of $(k-2) \times (k-2)$, and of the subsets of $k-2$. For each $p \in \mathcal{P}$, let $f(0)$ code the instances of equality among $y_0, \ldots, y_{k-2}$, and for $1 \le m < \omega$, let $f(m)$ code which subset of $\{y_0, \ldots, y_{k-2}\}$ has the m-th predicate as part of their type. (Alternately, we could have enumerated the predicates with different variables $Q_0(y_0)$, $Q_0(y_1)$, ..., and let f take values in $\{0, 1\}$.)

Now, if we preserve initial segments of f, we clearly hold constant the types of the parameters up to some level k in our hard-coded tree. Lemma 2.9 tells us that m exists as a function of s, as desired.

Unless $\mathcal{H}$ is very uncomplicated (for example, cliques all the way up) the theory will normally be in case $(a)(1)$ of Remark 3.3. For a question relating to the pseudo-nfcp, see [8, 3.1.5]. □

4. A separation via flexibility

The theories built above are simple rank one (Lemma 2.11 above), and thus they are low. In this section, we consider flexible ultrafilters, those which Kunen called "OK", which are necessary to saturate any nonlow theory in Keisler's order (see [3]).

Definition 4.1. Recall that the ultrafilter $\mathcal{D}$ on I, $|I| = \lambda$, is *flexible* if it has a regularizing family below any nonstandard integer, that is, for every sequence of natural numbers $\langle n_i : i \in I\rangle$ such that $\prod_{i\in I} n_i/\mathcal{D} > \aleph_0$, there is $\{X_\alpha : \alpha < \lambda\} \subseteq \mathcal{D}$ such that for all $i \in I$,

$$|\{\alpha < \lambda : i \in X_\alpha\}| \leq n_i.$$

Definition 4.2. Recall that a necessary and sufficient condition for a regular ultrafilter $\mathcal{D}$ on I, $|I| = \lambda$, to be *good for the random graph* is that for any infinite M and any $A, B \subseteq M^I/\mathcal{D}$ such that $|A| + |B| \leq \lambda$ and $A \cap B = \varnothing$, there is an internal predicate P such that $A \subseteq P$, whereas $B \cap P = \varnothing$.

Theorem 4.3. *Suppose $\mathcal{D}$ is a regular ultrafilter on I, $|I| = \lambda$, which is flexible and good for the random graph. Suppose (T, φ) has the pseudo-nfcp and $M \models T$. Then $M^I/\mathcal{D}$ is λ^+-saturated for positive φ-types.*

Proof. Let $M \models T$ and let $N = M^I/\mathcal{D}$. Consider a positive φ-type $p(x)$, where $\varphi = \varphi(\bar{x}, \bar{y})$. Enumerate the type as $\langle \varphi_\alpha(\bar{x}, \bar{a}_\alpha) : \alpha < \lambda\rangle$. Fix $i_* = \langle i_t : t \in I\rangle/\mathcal{D}$ a nonstandard integer (so that "max" will be well defined). For a finite tuple $\bar{a}$ from N, let $f_{\bar{a}}$ mean $f_{\mathrm{tp}(\bar{a},\varnothing,N)}$ and given in addition an index $t \in I$, let $f_{\bar{a}[t]}$ mean $f_{\mathrm{tp}(\bar{a}[t],\varnothing,M)}$. For each $\alpha < \lambda$ and each $t \in I$ (i.e., for each formula and each index), define

- $n(\alpha, t)$ to be the largest $n \leq i_t$ such that for all $\ell < m$, the type of $\bar{a}_\alpha[t]$ aligns with that of $\bar{a}_\alpha$ up to level n as measured by f, that is,
$$n(\alpha, t) := \max\{n \leq i_t : f_{\bar{a}_\alpha[t]} \restriction n = f_{\bar{a}_\alpha} \restriction n\}.$$
- $s(\alpha, t) := G(n(\alpha, t))$, using the notation of Remark 3.3.

The first is well defined since the condition is trivially true for 0. By Łos' theorem, since the f's reflect formulas for each $n < \omega$ and each $\alpha < \lambda$,

$$\{t \in I : n < n(\alpha, t)\} \in \mathcal{D}.$$

Hence, for each $\alpha < \lambda$, $n_\alpha := \prod_t n(\alpha, t)/\mathcal{D}$ is a nonstandard integer. It follows from Remark 3.3(b)($\star$) that for each $\alpha < \lambda$, $s_\alpha := \prod_t s(\alpha, t)/\mathcal{D}$ is either "∞" on a large set, or a nonstandard integer.

Since $\mathcal{D}$ is good for the random graph, $\mathrm{lcf}(\omega, \mathcal{D}) \geq \lambda^+$, so there is a nonstandard integer $\boldsymbol{s} = \langle \boldsymbol{s}[t] : t \in I \rangle / \mathcal{D}$ such that for each $\alpha < \lambda$, $\boldsymbol{s} < s_\alpha \bmod \mathcal{D}$. Since $\mathcal{D}$ is flexible and $\boldsymbol{s}$ is a nonstandard integer, we may choose $\{X_\alpha : \alpha < \lambda\} \subseteq \mathcal{D}$ regularizing $\mathcal{D}$ and with the property that for each $t \in I$,

$$|\{\alpha < \lambda : t \in X_\alpha\}| \leq \boldsymbol{s}[t].$$

Define a map $d : [\lambda]^1 \to \mathcal{D}$ by

$$\{\alpha\} \mapsto \{t \in I : \boldsymbol{s}[t] < s(\alpha, t)\} \cap X_\alpha.$$

That is, we assign α to an index set where we can be sure that the type of each $\bar{a}_\alpha[t]$ is "correct" up to the level needed to handle $s(\alpha, t) = G(n(\alpha, t))$ instances, thus a fortiori $\boldsymbol{s}[t]$ instances. The intersection with X_α ensures, for each $t \in I$, the set $U(t) := \{\alpha : t \in d(\{\alpha\})\}$ of instances assigned to index t has size $\leq \boldsymbol{s}[t]$.

Now for each $t \in I$, in the ultrapower N, $\{\varphi(\bar{x}, \bar{a}_\alpha) : \alpha \in U_t\}$ is a set of no more than $\boldsymbol{s}[t]$ positive instances of φ, and by definition is a partial type. Also by our definition, for each α, and in particular for each $\alpha \in U_t$,

$$f_{\bar{a}_\alpha} \restriction n(\alpha, t) = f_{\bar{a}_\alpha[t]} \restriction n(\alpha, t).$$

It follows that $\{\varphi(\bar{x}, \bar{a}_\alpha[t]) : \alpha \in U_t\}$ remains a partial type in the index model M. So we can realize the type at each index under this distribution, and thus in the ultrapower N. $\square$

Corollary 4.4. *If T is a theory from Section 1, $M \models T$, and $\mathcal{D}$ is a regular ultrafilter on I, $|I| = \lambda$, which is flexible and good for the random graph, then $M^I/\mathcal{D}$ is λ^+-saturated.*

Proof. We argue almost identically to [7, Definition 4.7, Claim 4.8, Fact 5.2 and Conclusion 5.7] (changing just the arity of the edge relation, and eliminating the bipartition from the case of graphs) that in regular ultrapowers which are good for the theory of the random graph, for λ^+-saturation it suffices to consider partial types of the form

$$p(x) = \{Q_\nu(x)\} \cup \{R(x, \bar{a}) : \bar{a} \in {}^{k-2}A\}$$

for $\lg n(\nu) < \omega$. (Briefly, those definitions and claims note that any regular ultrapower has a certain weak saturation, for instance leaves are large, and instances of equality in types can be safely ignored. Now use quantifier elimination to get a simple normal form for types by specifying the leaf of x, a set of tuples it connects to, and a disjoint set of tuples it does not connect to. Since saturation of ultrapowers reduces to saturation of φ-types, it is sufficient to deal with only a finite amount of information on the leaf of x. Finally, since "goodness for the random graph" allows us to internally separate sets of size $\leq \lambda$, it suffices to handle the positive part of the type.) $\square$

Definition 4.5. For the purposes of the next corollaries, call a theory T a pseudo-nfcp theory if there is a set Σ of formulas of the language such that

(a) (T, φ) has the pseudo-nfcp for each $\varphi \in \Sigma$, and

(b) given any regular ultrafilter $\mathcal{D}$ over λ and $M \models T$, whether $M^\lambda/\mathcal{D}$ is λ^+-saturated depends only on λ^+-saturation for positive φ-types for $\varphi \in \Sigma$.

Corollary 4.6. *Let T be a pseudo-nfcp theory, and let $\trianglelefteq$ denote Keisler's order.*

(a) *Let T_* be any nonlow simple theory. Then $T \trianglelefteq T_*$.*

(b) $T \trianglelefteq T_{\text{feq}}$.

Thus, if T is a pseudo-nfcp theory and T_* is any nonlow or nonsimple theory, $T \trianglelefteq T_*$. In particular, this is true for all the theories of Section 1 above.

Proof. Any regular ultrafilter on $\lambda \geq \aleph_0$ which is good for some unstable theory is necessarily good for the random graph, as the random graph is the $\trianglelefteq$-minimum unstable theory. Any regular ultrafilter which is good for T_{feq} is flexible [3, Lemma 8.8], and indeed any regular ultrafilter $\mathcal{D}$ which is good for some nonlow simple theory is flexible [3, Lemma 8.7]. The last line of the corollary now follows from the fact that T_{feq} is the Keisler-minimum nonsimple theory [4, Theorem 13.1] (as T_{feq} is minimum among theories with TP_2, whereas SOP_2 implies maximality). □

Discussion 4.7. The current instances of incomparability in Keisler's order mostly use one of two main ideas. The first is to say on one hand, changing the distance in the alephs between λ and some smaller μ (the size of a maximal antichain in a certain Boolean algebra used in building the ultrafilter) affects for which values of k the theories $T_{k+1,k}$ are saturated, and on the other, the "canonical simple nonlow theory" (see appendix to [5]) requires the ultrafilter to be flexible; under large cardinal assumptions, these two indicators can be varied independently; see [10; 5]. In ZFC, this phenomenon can be scaled down to see an incomparability between the $T_{k+1,k}$'s and a certain theory based on trees, which is low [6]. A second, much larger scale of incomparability was produced in [7], with continuum many simple rank one theories, the graph precursors of the hypergraph theories built here. As this discussion suggests, and as the proofs of this section show, once the ultrafilter becomes flexible, the noise of any differences in the present theories is drowned out by the huge power of the regularizing families available. Do there exist incomparable simple nonlow theories? Is incomparability mainly visible in the absence of forking?

We also record that, as an interesting immediate consequence of earlier arguments [10; 5], the theories built in Section 2 are (assuming a large cardinal) distinguishable in Keisler's order from the theories $T_{k+1,k}$, the higher analogues of the triangle-free random graph from [1]. That is:

Conclusion 4.8. *Assuming a supercompact cardinal, for arbitrarily large λ and any $\ell < \omega$ there is a regular ultrafilter $\mathcal{D}$ on λ which is flexible and good for the random graph,* ***thus*** *good for theories of Section 2, but not good for $T_{k+1,k}$ for any $2 \leq k < \ell$.*

Proof. Claim 10.32 in [5] gives the existence of the needed ultrafilter and in clause (a) shows it is not good for $T_{k+1,k}$ for $k < \ell$. Claim 10.30 in [5] shows this ultrafilter is flexible and good for the random graph. So by Theorem 4.3 above it can handle the theories of Section 2. □

References

[1] E. Hrushovski, "Pseudo-finite fields and related structures", pp. 151–212 in *Model theory and applications* (Ravello, Italy), edited by L. Bélair et al., Quad. Mat. **11**, Aracne, Rome, 2002. MR Zbl

[2] H. J. Keisler, "Ultraproducts which are not saturated", *J. Symbolic Logic* **32** (1967), 23–46. MR Zbl

[3] M. Malliaris, "Hypergraph sequences as a tool for saturation of ultrapowers", *J. Symbolic Logic* **77**:1 (2012), 195–223. MR Zbl

[4] M. Malliaris and S. Shelah, "Cofinality spectrum theorems in model theory, set theory, and general topology", *J. Amer. Math. Soc.* **29**:1 (2016), 237–297. MR Zbl

[5] M. Malliaris and S. Shelah, "A new look at interpretability and saturation", *Ann. Pure Appl. Logic* **170**:5 (2019), 642–671. MR Zbl

[6] M. Malliaris and S. Shelah, "An example of a new simple theory", pp. 121–151 in *Trends in set theory* (Piscataway, New Jersey), edited by S. Coskey and G. Sargsyan, Contemp. Math. **752**, Amer. Math. Soc., Providence, RI, 2020. MR Zbl

[7] M. Malliaris and S. Shelah, "Keisler's order is not simple (and simple theories may not be either)", *Adv. Math.* **392** (2021), art. id. 108036. MR Zbl

[8] M. Malliaris and S. Shelah, "Some simple theories from a Boolean algebra point of view", *Ann. Pure Appl. Logic* **175**:1 (2024), art. id. 103345. MR Zbl

[9] S. Shelah, *Classification theory and the number of nonisomorphic models*, Studies in Logic and the Foundations of Mathematics **92**, North-Holland, Amsterdam, 1978. MR Zbl

[10] D. Ulrich, "Keisler's order is not linear, assuming a supercompact", *J. Symb. Log.* **83**:2 (2018), 634–641. MR Zbl

Received 31 Jan 2022. Revised 15 Jan 2023.

MARYANTHE MALLIARIS:

mem@math.uchicago.edu

Department of Mathematics, University of Chicago, Chicago, IL, United States

SAHARON SHELAH:

shelah@math.huji.ac.il

Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Jerusalem, Israel

and

Department of Mathematics, Rutgers, The State University of New Jersey, Piscataway, NJ, United States

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.465

How I got to like graph polynomials

Johann A. Makowsky

For Boris Zil'ber on his 75th birthday

I trace the roots of my collaboration with Boris Zil'ber, which combines categoricity theory, finite model theory, algorithmics, and combinatorics.

1. Introduction and dedication

Boris Zil'ber played a crucial role in my work on graph polynomials. Some of my work that he inspired and in which he contributed, is summarized in Kotek et al. [2011]. A preprint was posted as Makowsky and Zil'ber [2006] and a conference paper was published as Kotek et al. [2008]. These are our only jointly published papers. Since then, a general framework for studying graph polynomials has slowly evolved. It bears witness to the impact of Boris on my own work. In this paper, I will describe how I got to like graph polynomials. Boris and I both started our scientific career in model theory. Boris pursued his highly influential work in various directions of infinite model theory. My path towards graph polynomials took a detour into the foundations of computer science, only to lead me back to model theoretic methods in finite combinatorics. I describe here how, step-by-step, I ended up discussing graph polynomials with Boris. Some of those steps owe a lot to serendipity, as others were triggered by natural questions arising from previous steps. These steps are described in Sections 2–6. Sections 7–8 describe some of the original ideas underlying the model-theoretic approach to graph polynomials. Section 9 summarizes where this encounter with Boris has led me. Ultimately, it looks as if Boris' influence on my path was inevitable, but only in retrospect. Meeting Boris in Oxford was a chance encounter with unexpected consequences. I would like to thank Boris for leading me to a fruitful new research area. Happy birthday, and many years of productive mathematics to come, till 120.

I would like to thank J. Kirby for various suggestions on how to improve the paper.
MSC2020: 05C13, 05C35, 05C31, 68Q99.
Keywords: categoricity, parametrized complexity, graph algorithms, graph polynomials.

2. Morley's 1965 paper

My first attempt to tackle open problems in model theory was a consequence of reading Morley's fundamental paper [1965] on categoricity in power, in the *undergraduate seminar* in mathematical logic at ETH Zürich in 1969. The seminar was held by Specker and Läuchli, and regularly attended by the then still very lucid octogenarian Bernays.

Building on earlier work by Mostowski, Ehrenfeucht and Vaught, Morley proved in 1965 a *truly deep* theorem in model theory:

Theorem 1 (Morley's theorem). *Let T be a first order theory and assume T has no finite models and is κ-categorical for* ***some*** *uncountable κ. Then T is κ-categorical for* ***every*** *uncountable κ.*

More importantly, even, the paper ended with a *list of questions* and many logicians and mathematicians were attracted by these. Among them I remember Baldwin and Lachlan, Ressaire, Lascar, Makkai, Harnik, Shelah, Zil'ber, Taĭtslin and his school (see Taĭtslin [1970]) and myself. In my MSc thesis from 1971 [Makowsky 1974], I managed to prove the following:

Theorem 2. (i) *A first order theory T which is $\aleph_0$-categorical and strongly minimal (hence categorical in all infinite κ) cannot be finitely axiomatizable.*

(ii) *There is a finitely axiomatizable complete first order theory T without finite models which is superstable.*

(iii) *If there is an* ***infinite, finitely presentable group G with only finitely many conjugacy classes****, there is also a complete finitely axiomatizable $\aleph_1$-categorical theory T_G without finite models.*

After I finished my MSc thesis, Specker drew my attention to a Soviet paper by Mustafin and Taimanov [1970], and as a result of this, I started a correspondence with Taimanov. Before 1985, there were very few authors citing my work, among them Ahlbrandt (a PhD-student of Baldwin), Rothmaler, Tuschik (from the German Democratic Republic), Zil'ber, Peretjat'kin and Slissenko from the Soviet Union. Boris was one of the first to notice and cite my work on categoricity. I soon realized that I could not make any further progress on these questions. I had no new ideas, and competition was overwhelming. Shelah's sequence of papers inspired by these open questions led many young researchers to abandon this direction of research in model theory. The finite axiomatizability questions were finally solved by Peretjat'kin [1980] and Zil'ber [1981]. Peretjat'kin constructed a finitely axiomatizable theory categorical in $\aleph_1$ but not in $\aleph_0$. Zil'ber showed that no finitely axiomatizable totally categorical first order theory exists. An alternative proof of this was given by Cherlin et al. [1985].

My first acquaintance with Boris Zil'ber happened via the literature. But our paths diverged (not in the yellow wood), and we did meet personally, but not very often.

3. From abstract model theory to computer science and graph algorithms

After leaving Morley-type model theory, I first worked in abstract model theory, and then in theoretical computer science. In computer science I dealt with the foundations of database theory and logic programming, which led me to finite model theory. My main tools from model theory were pebble games and the Feferman–Vaught theorem and its generalizations. Around this time I met Courcelle and became aware of the Robertson–Seymour theorems and their applications to graph algorithms described by Fellows [1989]. But it was Courcelle [1992] who first observed that logical methods would give even more applications, Courcelle's work on the monadic second order theory of graphs is summarized in the monumental monograph from 2012 by Courcelle and Engelfriet [2012].

Let $d(G)$ be a graph parameter and P be a graph property. If deciding whether a graph G on n vertices with $d(G) = t$ is in P can be done in time $c(t) \cdot n^s$ for some fixed s which does not depend on $d(G)$, nor on the number of vertices of G, we say that *P is Fixed Parameter Tractable* (FPT). This concept was introduced by Downey and Fellows [2013].

Theorem 3 (Courcelle, 1992). *Let C be a class of finite graphs of tree-width at most t, and let P be a graph property definable in monadic second order logic* MSOL. *Then checking whether a graph G is in C with n vertices is in P is in* FPT, *in fact*, *it can be solved in linear time $c(t)n$.*

In the mid-1990s, two students were about to change my research dramatically. My former master's student, Udi Rotics, returned from his experience in industry. His MSc thesis dealt with the logical foundation of databases. However, now he wanted to work on a PhD in *graph algorithms* but *without involving logic*. He proposed to extend the notion of *tree-width* of a class of graphs as a graph parameter in order to get a new width parameter which one can use for *fixed parameter tractability*. Finally, *but still using logic* (MSOL), we came up with a notion roughly equivalent to *clique-width*, introduced recently by Courcelle and Olariu [2000]. This led to my intensive collaboration with Courcelle and Rotics [Courcelle et al. 1998; 2000; 2001]. In my own paper [Makowsky 2004], I examine the algorithmic uses of the Feferman–Vaught theorem for fixed parameter tractability. Applications of my work with Courcelle and Rotics are well summarized in Downey and Fellows [2013].

In 1996, I started to supervise an immigrant student from the former USSR, Gregory Kogan, who wanted to work on the complexity of computing the *permanent*. He came with impressive letters of recommendation. He had some spectacular partial results for computing permanents of matrices over a field of characteristic 3. He was a virtuoso in combinatorial linear algebra. Unfortunately, he dropped out before finishing his PhD. Kaminski and I wrote up his results, published under his name alone as Kogan [1996].

4. Computing permanents

I first came across the problem of computing the permanent at Specker's 60th birthday conference in 1980. The permanent of an $(n \times n)$-matrix $A = (A_{i,j})$ is given as

$$\operatorname{per}(A) = \sum_{s:[n]\to[n]} \prod_{i\in[n]} A_{i,s(i)},$$

where s ranges over all permutations of $[n]$.

The complexity class $\sharp\mathbf{P}$ is the polynomial time counting class.

The class of $\sharp\mathbf{P}$ consists of function problems of the form "compute $f(x)$", where f is the number of accepting paths of a nondeterministic Turing machine running in polynomial time. Unlike most well-known complexity classes, it is not a class of decision problems but a class of function problems. The most difficult representative problems of this class are $\sharp\mathbf{P}$-complete. Counting the number of satisfying assignments for a formula of propositional logic is $\sharp\mathbf{P}$-complete. Typical examples would be described as follows: Let k be a fixed integer. Given an input graph G on n vertices, compute the number of proper k-colorings of G. For $k = 1, 2$, this can be computed in polynomial time, but for $k \geq 3$, this is $\sharp\mathbf{P}$-complete with respect to **P**-time reductions. In general, $\sharp\mathbf{P}$ lies between the polynomial hierarchy **PH** and **PSpace**; see Papadimitriou [1994].

Valiant's complexity classes **VP** and **VNP** are the analogues of **P** and **NP** in Valiant's model of algebraic computation. Bürgisser's book [2000] is entirely dedicated to this model of computation. It is still open whether $\mathbf{P} = \mathbf{NP}$, and also whether $\mathbf{VP} = \mathbf{VNP}$.

Valiant presented the complexity classes **VP** and **VNP** at Specker's 60th birthday conference.

Theorem 4 [Valiant 1979]. *Computing the permanent of a* $\{0, 1\}$*-matrix is hard in the following sense*:

(i) *It is* $\sharp\mathbf{P}$*-complete in the Turing model of computation*, *and*

(ii) **VNP***-complete in Valiant's algebraic model of computation.*

Valiant [1982] published in the proceedings of Specker's 60th birthday conference in 1980. Bürgisser's monograph [2000] explores Valiant's approach to algebraic complexity further.

Kogan studied the complexity of computing the permanent over fields of characteristic 3 for matrices M with $\mathrm{rank}(MM^{\mathrm{tr}} - \mathbf{1}) = a$. He showed that for $a \leq 1$, this is easy; and for $a \geq 2$, this is hard.

I wanted to use results from Courcelle et al. [2001] to prove something about permanents Kogan could not prove. I looked at adjacency matrices of graphs of fixed tree-width t. Barvinok [1996] also studied the complexity of computing the permanent for special matrices. He looked at matrices of fixed rank r. Our results were:

Theorem 5 (Barvinok, 1996). *Let $\mathcal{M}_r$ be the set of real matrices of fixed rank r. There is a polynomial time algorithm $\mathcal{A}_r$ which computes* $\mathrm{per}(A)$ *for every $A \in \mathcal{M}_r$*

Theorem 6 (JAM, 1996). *Let $\mathcal{T}_w$ be the set of adjacency matrices of graphs of tree-width at most w. There is a polynomial time algorithm $\mathcal{A}_w$ which computes* $\mathrm{per}(A)$ *for every $A \in \mathcal{T}_w$.*

The two theorems are incomparable. There are matrices of tree-width t and arbitrary large rank, and there are matrices of rank r and arbitrary large tree-width.

However, I realized that the proof of my theorem had nothing to do with permanents. It was much more general and really worked quite generally. It only depended on some *logical restrictions* for polynomials in indeterminates given by the entries of the matrix. If the matrix $A = A_G$ is the adjacency matrix of a graph G where the nonzero entries are x, the permanent $\mathrm{per}(A_G)$ can be viewed as a graph polynomial in the indeterminate x. Alas, at that time I had no clue how to find many interesting examples.

5. From knot polynomials to graph polynomials

During a sabbatical at ETH in Zürich I met Turaev, who, among other things, is an expert in knot theory. I showed him my Theorem 6. He suggested I should try to prove the same for the *Jones polynomial* from knot theory. So I studied knot theory intensively for a few months. While visiting the Fields Institute in 1999, I attended a lecture by Mighton[1] who lectured about the Jones polynomial for series-parallel knot diagrams; see his PhD thesis [Mighton 2000]. He showed that, in this case, the Jones polynomials is computable in polynomial time. Series-parallel graphs are exactly the graphs of tree-width 2. It seemed reasonable that the same would hold for graphs of tree-width k. Indeed, after quite an effort I proved the following in [Makowsky 2001; 2005]:

Theorem 7. *Assume K is a knot with knot diagram D_k of tree-width k. Then evaluating the Jones polynomial $J(D_k; a, b)$ for fixed complex numbers $a, b \in \mathbb{C}$ and D_k with n vertices is in* **FPT** *with parameter k.*

[1] John Mighton is a Canadian mathematician, author and playwright.

Jaeger et al. [1990] showed that, without the assumption on tree-width, evaluating the Jones polynomial is $\sharp\mathbf{P}$-complete for almost all $a, b \in \mathbb{C}$. Lotz and Makowsky [2004] analyze the complexity of the Jones polynomial in Valiant's model of computation.

However, again the proof seemed to work for other graph polynomials as well, among them the *Tutte polynomial*, *chromatic polynomial*, *characteristic polynomial*, *matching polynomial*. Univariate graph polynomials are graph invariants which take values in a polynomial ring, usually $\mathbb{Z}[X]$, $\mathbb{R}[X]$ or $\mathbb{C}[X]$. The univariate chromatic polynomial $\chi(G; k)$ of a graph counts the number of proper colorings of a graph with at most k colors. It was introduced by Birkhoff in 1912 in an unsuccessful attempt to prove the four color conjecture. The characteristic polynomial of a graph is the characteristic polynomial (in the sense of linear algebra) of the adjacency matrix A_G of the graph G; see the monographs by Chung [1997] and by Brouwer and Haemers [2012]. The coefficients of X^k of the matching polynomial count the number of k-matchings of a graph G; see Lovász and Plummer [2009]. Both, the characteristic and the matching polynomial, have found applications in *theoretical chemistry* as described by Trinajstić [1992]. There are also multivariate graph polynomials. The Tutte polynomial is a bivariate generalization of the chromatic polynomial. Both of them are widely studied; see Dong et al. [2005] and the handbook of the Tutte polynomial edited by Ellis-Monaghan and Moffatt [2022]. Other widely studied graph polynomials are listed in Makowsky [2008]. However, I had no idea, how to find *infinitely many natural and interesting examples*?

6. Boris, deus ex machina

In 2005, while attending CSL, the European Conference in Computer Science Logic in Oxford, I paid a visit to Boris Zil'ber, whom I knew and had met before due to our work on Morley's problem on finite axiomatizability of totally categorical theories. After a few friendly exchanges the following dialogue evolved:

Boris: What do you work on nowadays?

Me: Graph polynomials.

Boris: What polynomials?

It seemed Boris had never heard of graph polynomials. I gave him the standard examples (Tutte, chromatic, matching). He immediately saw them as examples which are interpretable in some totally categorical theory. I could not believe it.

We spent the next days together, verifying that all the known graph polynomials fit into Zil'ber's framework. It was indeed the case. We also produced generalizations of chromatic polynomials, some of which I later called *Harary polynomials* [Herscovici et al. 2021; 2020]. They are generalizations of the chromatic polynomial based on conditional colorings introduced in Harary [1985] in 1985.

Conditional colorings are defined using a graph property P. A P-coloring f of G with at most k colors is a function $f : V \to [k]$ such that for every color $j \in [k]$, the set $f^{-1}(j)$ induces a graph in P. Conditional colorings are studied in the literature, e.g., by Brown and Corneil [1987], mostly in the context of extremal graph theory. However, nobody wrote about the fact that counting the number of such colorings with at most k colors defines a polynomial in k. The so called Harary polynomial $\chi_P(G; k)$ counts the number of P-colorings of G with at most k colors.

7. Why is the chromatic polynomial of a graph a polynomial?

Let $G = (V, E)$ be a graph. A proper coloring of G with at most k colors is a function $f : V \to [k]$ such $f(v) = f(v')$ implies that $\neg E(v, v')$. We think of $[k]$ as a set of colors. In other words, if two vertices have the same color they are not adjacent. We denote by $\chi(G; k)$ the number of proper colorings of G with at most k colors. Birkhoff's proof that $\chi(G; k)$ is a polynomial in $\mathbb{Z}[k]$ uses deletion and contraction of edges. Let $e = (u, v)$ be an edge of G. G_{-e} is the graph $G_{-e} = (V, E - \{(u, v)\})$ where e is deleted from E. $G_{/e}$ is the graph $G_{/e} = (V_{/e}, E|_{V_{/e}})$ where e is contracted to a single vertex to form $V_{/e}$ and e is omitted from E. f is a proper coloring of G_{-e} if either it is a proper coloring of G and $f(u) \neq f(v)$ or it is a proper coloring of $G_{/e}$ and $f(u) = f(v)$. Furthermore, $\chi(G; k)$ is multiplicative, i.e., if G is the disjoint union of G_1 and G_2, then

$$\chi(G_1 \sqcup G_2; k) = \chi(G_1; k) \cdot \chi(G_2; k).$$

Let E_n be the edgeless graph with n vertices and$E = \emptyset$. We have $\chi(E_n; k) = k^n$, and

$$\chi(G_{-e}; k) = \chi(G; k) + \chi(G_{/e}; k).$$

By showing that one can compute $\chi(G; k)$ by successively removing edges, and this is independent of the order of the edges, one concludes that $\chi(G; k)$ is a polynomial in k. The disadvantage of this elegant proof is, that it does not generalize.

Another way of proving that $\chi(G; k)$ is a polynomial in k is by noting that for graphs on n vertices, we have

$$\chi(G; k) = \sum_{i=1}^{n} c_i(G) k_{(i)},$$

where the coefficient $c_i(G)$ is the number of proper colorings of G with exactly i colors and

$$k_{(i)} = k \cdot (k-1) \cdots (k-i+1) = \prod_{i=0}^{i} (k-i) = \binom{k}{i} \cdot i!$$

the falling factorial. Note that $\binom{k}{i} = 0$ for $i > k$. As $k_{(i)}$ is a polynomial in k and $\chi(G; k)$ is a sum of n polynomials in k, the result follows. However, this proof *does generalize*, and it works for all Harary polynomials.

Later I discussed Zil'ber's view of graph polynomials with Blass. We noted that for most of the graph invariants from the literature, proving that they were polynomial invariants via totally categorical theories was an overkill. This led me to formulate a considerably simplified approach, which indeed covered all the know examples of graph polynomials in the literature. This approach is a simplification of Boris' proof. It generalizes also to other types of graph polynomials such as the bivariate Tutte polynomial and the trivariate edge elimination polynomial from Averbouch et al. [2010]. More intrinsic examples are also discussed in Makowsky and Zil'ber [2006] and Kotek et al. [2011]. However, the polynomial graph invariants hidden in totally categorical theories are the most general graph invariants which are definable in second order logic SOL, and even in higher order logic HOL, over the graph G; see Makowsky and Zil'ber [2006, Corollary C and Theorem 3.15]. Furthermore, it applies to HOL-definable polynomial invariants over arbitrary finite first order structures for finite vocabularies, rather than just to graphs.

8. The model-theoretic approach to the chromatic polynomial

The way Boris looked at the chromatic polynomial was even more general. Given a graph G, Boris had in mind an infinite first order structure $\mathfrak{M}(G)$ with universe M, and a formula $\phi(x)$ such that:

(i) The first order theory $T(\mathfrak{M}(G))$ of $\mathfrak{M}(G)$ is totally categorical and strongly minimal with a strongly minimal infinite set X of indiscernibles.

(ii) The first order theory $T(\mathfrak{M}(G))$ has the finite model property, i.e., the algebraic closure $\mathrm{acl}(Y)$ in $\mathfrak{M}(G)$ of a finite subset $Y \subset X$ is finite.

(iii) $\mathfrak{M}(G) \models \phi(x)$ if and only if x is a proper coloring of G.

Such theories were at the heart of his work [Zil'ber 1993]. From Zil'ber's analysis of totally categorical theories, we get in the spirit of [Zil'ber 1993, Theorem 1.5.5].

Theorem 8. *Let $\mathfrak{M}(G)$ and X as above. For every finite set $Y \subset X$ of cardinality k the cardinality of the set*

$$\{x \in \mathrm{acl}(Y) : \mathfrak{M}(G) \models \phi(x)\}$$

is a polynomial in k.

In the case of the chromatic polynomial this looks as follows:

(i) Let $G = (V, E)$ be a finite graph, with $|V| = n$.

(ii) Let $\mathfrak{M}(G) = (V, X; E, \bar{v})$ be a 2-sorted language with sorts V, X, a binary relation E on V for the edge relation, and n constant symbols $v_1, \dots, v_n$ of sort V.

(iii) The describing axioms state that (V, E) is exactly the finite graph we started with, and the vertices are exactly the v_i. Then a model of the axioms is specified up to isomorphism by the cardinality of X.

(iv) If we add axioms stating that X is infinite then the first order theory $T(\mathfrak{M})$ is totally categorical.

(v) We also get finite models M_k, where $|X| = k$ for each natural number k, and they are algebraically closed subsets of the infinite model $\mathfrak{M}$.

(vi) We regard X as a set of colors, and we identify X^n with the colorings of vertices, that is, the set of functions $V \to X$, by identifying $f = (x_1, \ldots, x_n) \in X^n$ with the function $f(v_i) = x_i$.

(vii) The map f is a *proper vertex coloring* if any two adjacent vertices have different colors. So the set of proper vertex colorings is defined as a subset of X^n by the formula $\phi(\bar{x})$ given by

$$\phi(\bar{x}) : \bigwedge_{\{(i,j):E(v_i,v_j)\}} x_i \neq x_j.$$

(viii) The chromatic polynomial for the graph G is $\chi(G; k) = |\phi(M_k)|$.

Boris also showed me at our first encounter how the bivariate matching polynomial and the Tutte polynomial can be cast in this framework. For the characteristic polynomial $p(G; x)$ the situation is a bit more complicated, because its original definition uses the characteristic polynomial of the adjacency matrix $A(G)$ of G. However, there exists a purely graph theoretic description of the coefficients of $p(G; x)$ by Godsil [1993], which allows to cast $p(G; x)$ into this framework. We note that it may be unexpectedly tricky to put a graph invariant into Boris' framework, even if one already knows that it is a polynomial invariant.

The most general version of this can be found in Kotek et al. [2011, Section 8]. Using this method, any multivariate polynomial graph invariant deinable in HOL can be captured in this way. In the last ten years, Nešetřil and his various collaborators (Goodall, Garijo and Ossona de Mendez) were exploring various ways to define such graph invariants. However, they did not reach the same generality; see [Garijo et al. 2011; 2016; 2016]. The potential of the general approach as described in [Kotek et al. 2011, Section 8] still has not been explored in depth. It seems that its abstract generality makes it difficult for the combinatorics community to see through this construction. On the other side, model theorists seemingly are not interested in combinatorial applications of model theory. Exceptions may be in extremal combinatorics, as initiated by Razborov [2007; 2013] and surveyed by Coregliano and Razborov [2020]. Another direction is counting the number $S_P(n)$ of graphs on n vertices in a hereditary graph property P, initiated by Scheinerman and Zito [1994] and further pursued by Balogh et al. [2000] and Laskowski and Terry [2022].

9. Towards a general theory of graph polynomials

For the last 20 years, I have been studying graph polynomials [Makowsky 2006; 2008], aiming to understand what they have in common. I discovered that:

- Graph polynomials can be studied to obtain *information on graphs*. As an example: Evaluations of the Tutte polynomial encodes many graph invariants.
- Graph polynomials can be studied as *polynomials indexed by graphs*. As an example: The acyclic matching polynomial of paths, cycles, complete graphs, and complete bipartite graphs are the Chebyshev polynomials of the second and first kinds, Hermite polynomials, and Laguerre polynomials, respectively.
- Two graph polynomials have the same *distinctive power* if they do not distinguish between the same two graphs.
- General theorems about graph polynomials often can be formulated as meta-theorems; see [Makowsky 2024].

With my various collaborators I managed to create a new field in graph theory with two Dagstuhl Seminars (16241, 19401), two MATRIX Institute programs, two special sessions at AMS meetings and one SIAM mini-symposium.

Without Boris Zil'ber's eye opener I would not have pursued this line of research as far as I did. Thank you, Boris!

References

[Averbouch et al. 2010] I. Averbouch, B. Godlin, and J. A. Makowsky, "An extension of the bivariate chromatic polynomial", *European J. Combin.* **31**:1 (2010), 1–17. MR Zbl

[Balogh et al. 2000] J. Balogh, B. Bollobás, and D. Weinreich, "The speed of hereditary properties of graphs", *J. Combin. Theory Ser. B* **79**:2 (2000), 131–156. MR Zbl

[Barvinok 1996] A. I. Barvinok, "Two algorithmic results for the traveling salesman problem", *Math. Oper. Res.* **21**:1 (1996), 65–84. MR Zbl

[Brouwer and Haemers 2012] A. E. Brouwer and W. H. Haemers, *Spectra of graphs*, Springer, New York, 2012. MR Zbl

[Brown and Corneil 1987] J. I. Brown and D. G. Corneil, "On generalized graph colorings", *J. Graph Theory* **11**:1 (1987), 87–99. MR Zbl

[Bürgisser 2000] P. Bürgisser, *Completeness and reduction in algebraic complexity theory*, Algorithms and Computation in Mathematics **7**, Springer, Berlin, 2000. MR Zbl

[Cherlin et al. 1985] G. Cherlin, L. Harrington, and A. H. Lachlan, "$\aleph_0$-categorical, $\aleph_0$-stable structures", *Ann. Pure Appl. Logic* **28**:2 (1985), 103–135. MR Zbl

[Chung 1997] F. R. K. Chung, *Spectral graph theory*, CBMS Regional Conference Series in Mathematics **92**, American Mathematical Society, Providence, RI, 1997. MR Zbl

[Coregliano and Razborov 2020] L. N. Coregliano and A. A. Razborov, "Semantic limits of dense combinatorial objects", *Uspekhi Mat. Nauk* **75**:4 (2020), 45–152. In Russian; translated in *Russian Math. Surveys* **75**:4 (2020), 627–723. MR Zbl

[Courcelle 1992] B. Courcelle, "The monadic second-order logic of graphs, III: tree-decompositions, minors and complexity issues", *RAIRO Inform. Théor. Appl.* **26**:3 (1992), 257–286. MR Zbl

[Courcelle and Engelfriet 2012] B. Courcelle and J. Engelfriet, *Graph structure and monadic second-order logic*, Encyclopedia of Mathematics and its Applications **138**, Cambridge Univ. Press, 2012. MR Zbl

[Courcelle and Olariu 2000] B. Courcelle and S. Olariu, "Upper bounds to the clique width of graphs", *Discrete Appl. Math.* **101**:1–3 (2000), 77–114. MR Zbl

[Courcelle et al. 1998] B. Courcelle, J. A. Makowsky, and U. Rotics, "Linear time solvable optimization problems on graphs of bounded clique width", pp. 1–16 in *Graph-theoretic concepts in computer science* (Smolenice Castle, 1998), edited by J. Hromkovič and O. Sýkora, Lecture Notes in Comput. Sci. **1517**, Springer, Berlin, 1998. MR Zbl

[Courcelle et al. 2000] B. Courcelle, J. A. Makowsky, and U. Rotics, "Linear time solvable optimization problems on graphs of bounded clique-width", *Theory Comput. Syst.* **33**:2 (2000), 125–150. MR Zbl

[Courcelle et al. 2001] B. Courcelle, J. A. Makowsky, and U. Rotics, "On the fixed parameter complexity of graph enumeration problems definable in monadic second-order logic", *Discrete Appl. Math.* **108**:1 (2001), 23–52. MR Zbl

[Dong et al. 2005] F. M. Dong, K. M. Koh, and K. L. Teo, *Chromatic polynomials and chromaticity of graphs*, World Scientific, Hackensack, NJ, 2005. MR Zbl

[Downey and Fellows 2013] R. G. Downey and M. R. Fellows, *Fundamentals of parameterized complexity*, Springer, London, 2013. MR Zbl

[Ellis-Monaghan and Moffatt 2022] J. A. Ellis-Monaghan and I. Moffatt, *Handbook of the Tutte polynomial and related topics*, CRC Press, Boca Raton, FL, 2022. Zbl

[Fellows 1989] M. R. Fellows, "The Robertson–Seymour theorems: a survey of applications", pp. 1–18 in *Graphs and algorithms* (Boulder, CO, 1987), edited by R. B. Richter, Contemp. Math. **89**, American Mathematical Society, Providence, RI, 1989. MR Zbl

[Garijo et al. 2011] D. Garijo, A. Goodall, and J. Nešetřil, "Distinguishing graphs by their left and right homomorphism profiles", *European J. Combin.* **32**:7 (2011), 1025–1053. MR Zbl

[Garijo et al. 2016] D. Garijo, A. Goodall, and J. Nešetřil, "Polynomial graph invariants from homomorphism numbers", *Discrete Math.* **339**:4 (2016), 1315–1328. MR Zbl

[Godsil 1993] C. D. Godsil, *Algebraic combinatorics*, Chapman and Hall, New York, 1993. MR Zbl

[Goodall et al. 2016] A. J. Goodall, J. Nešetřil, and P. Ossona de Mendez, "Strongly polynomial sequences as interpretations", *J. Appl. Log.* **18** (2016), 129–149. MR Zbl

[Harary 1985] F. Harary, "Conditional colorability in graphs", pp. 127–136 in *Graphs and applications* (Boulder, CO, 1982), edited by F. Harary and J. S. Maybee, Wiley, New York, 1985. MR Zbl

[Herscovici et al. 2020] O. Herscovici, J. A. Makowsky, and V. Rakita, "Harary polynomials", preprint, 2020. arXiv 2003.06250

[Herscovici et al. 2021] O. Herscovici, J. A. Makowsky, and V. Rakita, "Harary polynomials", *Enumer. Comb. Appl.* **1**:2 (2021), art. id. S2R13. MR Zbl

[Jaeger et al. 1990] F. Jaeger, D. L. Vertigan, and D. J. A. Welsh, "On the computational complexity of the Jones and Tutte polynomials", *Math. Proc. Cambridge Philos. Soc.* **108**:1 (1990), 35–53. MR Zbl

[Kogan 1996] G. Kogan, "Computing permanents over fields of characteristic 3: where and why it becomes difficult", pp. 108–114 in 37*th Annual Symposium on Foundations of Computer Science* (Burlington, VT, 1996), IEEE Computer Society, Los Alamitos, CA, 1996. MR Zbl

[Kotek et al. 2008] T. Kotek, J. A. Makowsky, and B. Zil'ber, "On counting generalized colorings", pp. 339–353 in *Computer science logic*, edited by M. Kaminski and S. Martini, Lecture Notes in Comput. Sci. **5213**, Springer, Berlin, 2008. MR Zbl

[Kotek et al. 2011] T. Kotek, J. A. Makowsky, and B. Zil'ber, "On counting generalized colorings", pp. 207–241 in *Model theoretic methods in finite combinatorics*, edited by M. Grohe and J. A. Makowsky, Contemp. Math. **558**, American Mathematical Society, Providence, RI, 2011. MR Zbl

[Laskowski and Terry 2022] M. C. Laskowski and C. A. Terry, "Jumps in speeds of hereditary properties in finite relational languages", *J. Combin. Theory Ser. B* **154** (2022), 93–135. MR Zbl

[Lotz and Makowsky 2004] M. Lotz and J. A. Makowsky, "On the algebraic complexity of some families of coloured Tutte polynomials", *Adv. in Appl. Math.* **32**:2 (2004), 327–349. MR Zbl

[Lovász and Plummer 2009] L. Lovász and M. D. Plummer, *Matching theory*, American Mathematical Society, Providence, RI, 2009. MR Zbl

[Makowsky 1974] J. A. Makowsky, "On some conjectures connected with complete sentences", *Fund. Math.* **81** (1974), 193–202. MR Zbl

[Makowsky 2001] J. A. Makowsky, "Colored Tutte polynomials and Kauffman brackets for graphs of bounded tree width", pp. 487–495 in *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms* (Washington, DC, 2001), SIAM, Philadelphia, PA, 2001. MR Zbl

[Makowsky 2004] J. A. Makowsky, "Algorithmic uses of the Feferman–Vaught theorem", *Ann. Pure Appl. Logic* **126**:1–3 (2004), 159–213. MR Zbl

[Makowsky 2005] J. A. Makowsky, "Coloured Tutte polynomials and Kauffman brackets for graphs of bounded tree width", *Discrete Appl. Math.* **145**:2 (2005), 276–290. MR Zbl

[Makowsky 2006] J. A. Makowsky, "From a zoo to a zoology: descriptive complexity for graph polynomials", pp. 330–341 in *Logical approaches to computational barriers* (Swansea, UK, 2006), Springer, Heidelberg, 2006. Zbl

[Makowsky 2008] J. A. Makowsky, "From a zoo to a zoology: towards a general theory of graph polynomials", *Theory Comput. Syst.* **43**:3-4 (2008), 542–562. MR Zbl

[Makowsky 2024] J. A. Makowsky, "Meta-theorems for graph polynomials", preprint, 2024. To appear in *MATRIX Annals*. arXiv 2405.02617

[Makowsky and Zil'ber 2006] J. Makowsky and B. Zil'ber, "Polynomial invariants of graphs and totally categorical theories", preprint, 2006, available at https://www.logique.jussieu.fr/modnet/Publications/Preprint%20server/papers/21. Zbl

[Mighton 2000] J. S. Mighton, *Knot theory on bipartite graphs*, Ph.D. thesis, University of Toronto, 2000, available at https://hdl.handle.net/1807/13488. Zbl

[Morley 1965] M. Morley, "Categoricity in power", *Trans. Amer. Math. Soc.* **114** (1965), 514–538. MR Zbl

[Mustafin and Taĭmanov 1970] T. G. Mustafin and A. D. Taĭmanov, "Countable models of theories which are categorical in power $\aleph_1$ but not in power $\aleph_0$", *Algebra i Logika* **9** (1970), 559–565. MR Zbl

[Papadimitriou 1994] C. H. Papadimitriou, *Computational complexity*, Addison-Wesley, Reading, MA, 1994. MR Zbl

[Peretjat'kin 1980] M. G. Peretjat'kin, "Example of an ω_1-categorical complete finitely axiomatizable theory", *Algebra i Logika* **19**:3 (1980), 314–347. MR Zbl

[Razborov 2007] A. A. Razborov, "Flag algebras", *J. Symbolic Logic* **72**:4 (2007), 1239–1282. MR Zbl

[Razborov 2013] A. A. Razborov, "Flag algebras: an interim report", pp. 207–232 in *The mathematics of Paul Erdős*, *II*, edited by J. N. Ronald L. Graham and S. Butler, Springer, New York, 2013. MR Zbl

[Scheinerman and Zito 1994] E. R. Scheinerman and J. Zito, "On the size of hereditary classes of graphs", *J. Combin. Theory Ser. B* **61**:1 (1994), 16–39. MR Zbl

[Taĭtslin 1970] M. A. Taĭtslin, Теория моделей, Publications of the Department of Algebra and Mathematical Logic of the Novosibirsk State University **6**, Novosibirsk. Gosudarstv. Univ., Novosibirsk, 1970. MR

[Trinajstić 1992] N. Trinajstić, *Chemical graph theory*, 2nd ed., CRC Press, Boca Raton, FL, 1992. MR Zbl

[Valiant 1979] L. G. Valiant, "The complexity of computing the permanent", *Theoret. Comput. Sci.* **8**:2 (1979), 189–201. MR Zbl

[Valiant 1982] L. G. Valiant, "Reducibility by algebraic projections", pp. 365–380 in *Logic and algorithmic* (Zurich, 1980), Monogr. Enseign. Math. **30**, Université de Genève, 1982. MR Zbl

[Zil'ber 1981] B. I. Zil'ber, "Totally categorical structures and combinatorial geometry", *Dokl. Akad. Nauk SSSR* **259**:5 (1981), 1039–1041. In Russian. MR Zbl

[Zil'ber 1993] B. Zil'ber, *Uncountably categorical theories*, Translations of Mathematical Monographs **117**, American Mathematical Society, Providence, RI, 1993. MR Zbl

Received 7 Feb 2023. Revised 5 Aug 2023.

JOHANN A. MAKOWSKY:

janos@cs.technion.ac.il
Faculty of Computer Science, Technion–Israel Institute of Technology Haifa, Israel

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.479

La conjecture d'algébricité, dans une perspective historique, et surtout modèle-théorique

Bruno Poizat

В статье описывается влияние гипотезы алгебраичности Черлина и Зильбера простых групп конечного ранга Морли, начиная с двух ее ишодных формулировок. В ней больше внимания уделяется их теоретико-модельным аспектам, чем алгебраическим. Излагаюстя истории, связанные со свойствами аддитивности ранга Морли и определениями групп конечкого ранга Морли по Боровику. Рассказывается о теореме о неразложимых множествах, о характеристике этих групп генерическими данными и о возможном распространении их свойств на структуры, более слабые, чем группы.

This paper describes the influence of the algebraicity conjecture of Cherlin and Zilber, concerning the simple groups of finite Morley rank, since its two original formulations. It insists on its model theoretic aspects more than on its algebraic aspects. It relates the history of the additivity properties of Morley rank and of the definition à la Borovik of groups of finite Morley rank. It accounts for the indecomposable sets theorem, the characterisation of these groups by generic data, and the possible extension of their properties to structures weaker than groups.

Cet article décrit l'influence de la conjecture d'algébricité de Zilber et de Cherlin, à propos des groupes simples de rang de Morley fini, depuis ses deux formulations originelles. Il insiste sur ses aspects plutôt modèle-théoriques qu'algébriques. On y fait l'historique des propriétés d'additivité du rang de Morley et de la définition à la Borovik des groupes de rang de Morley fini. On parle du théorème des indécomposables, de caractérisation des groupes par des données génériques, et de l'extension éventuelle de leurs propriétés à des structures plus faibles que des groupes.

Then spoke the king and said in Aramaic language: What! I hear that Zilber–Cherlin conjecture is open since fifty years! Who are these useless scholars who make conjectures instead of proving theorems? The thing is gone from me: if next morrow ye will not make the proof known unto me, ye shall be cut in pieces and your houses shall be made a dunghill!

Daniel 2.5, Unauthorized Version

MSC2020 : primary 03C45; secondary 03C60, 20E32, 20F11, 20F50.
Mots-clefs : groupes algébriques, groupes finis, groupes de rang de Morley fini, groupes superstables, groupes localement finis, conjecture de Cherlin–Zilber, espaces de symétries, symétrons.

1. Zilber et Cherlin

À la fin de son article aux Fundamenta, Boris Zilber [1977] pose quatre questions, que je reproduis ici accompagnées d'une traduction aussi littérale que possible, à l'intention de mes lectrices, et surtout de mes lecteurs, qui auraient du mal à apprécier les finesses de l'original.

(1) Существуют ли неабелевы связные слабо категоричные группы, неизоморфные алгебраическим группам над алгебраически замкнутым полем?

(1) Existe-t-il des groupes non abéliens connexes faiblement catégoriques qui ne soient isomorphes à un groupe algébrique sur un corps algébriquement clos ?

(2) Существуют ли простые категоричные группы, отличные от алгебраических над алгебраически замкнутым полем?

(2) Existe-t-il des groupes simples catégoriques distincts de ceux qui sont algébriques sur un corps algébriquement clos ?

(3) Можно ли группу $G = H_p + H_q$ интерпретировать в $\aleph_1$-категоричной теории? Здесь H_p, H_q — бесконечные абелены группы простых показателей p, q соответственно.

(3) Le groupe $G = H_p + H_q$ peut-il être interprété dans une théorie $\aleph_1$-catégorique ? Ici H_p, H_q sont des groupes abéliens infinis d'exposant premier p, q respectivement.

(4) Для любых ли двух $\aleph_1$-категоричных теорий T_1 и T_2 существует $\aleph_1$-категоричная теория T, в которой формулньо интерпретируются T_1 и T_2?

(4) Est-ce que, pour chaque paire de théories T_1 et T_2 $\aleph_1$-catégoriques, il existe une théorie $\aleph_1$-catégorique T avec des formules interprétant T_1 et T_2 ?

Il faut préciser que le cadre dans lequel travaille Zilber est celui d'une structure dont la théorie est $\aleph_1$-catégorique, et qu'il qualifie de *faiblement catégorique* ce qui y est interprétable. Ce contexte lui était familier depuis [Zilber 1974], où il donne une démonstration de la finitude du rang de Morley, tout en reconnaissant la priorité à [Baldwin 1973].

On sait maintenant que la réponse à la première question est positive. Par exemple, la fusion d'Ehud Hrushovski [1992] permet de définir sur un même ensemble fortement minimal deux corps algébriquement clos K et L de caractéristiques différentes, et on trouvera un contre-exemple en prenant le produit d'un groupe algébrique sur K et d'un groupe algébrique sur L ; ou encore en considérant le produit semi-direct du groupe additif du corps K par un de ses sous-groupes multiplicatifs M, où (K, M) est un corps vert de rang de Morley deux, dont la construction doit aussi beaucoup aux idées de Hrushovski [Poizat 2001a; Baudisch et al. 2009]. Un troisième exemple est constitué des groupes nilpotents d'Andreas Baudisch [1996], également construits par amalgame de Hrushovski.

La deuxième question est le premier avatar de la conjecture d'algébricité, qui reste ouverte aujourd'hui. On remarque un petit changement de vocabulaire : il est question d'un groupe simple catégorique, et non plus faiblement catégorique ; cela vient de ce que le théorème principal (теорема 5.2) de l'article déclare qu'un groupe simple faiblement catégorique est en fait $\aleph_1$-catégorique.

La réponse à la troisième question est également positive, car on peut fusionner les deux groupes.

Quant à la dernière, elle est bien entamée, mais reste encore du domaine de la conjecture. En effet, la fusion de Hrushovski permet d'amalgamer deux structures fortement minimales dénombrables en une troisième sous deux conditions : (i) elles sont toutes les deux ω-saturées, (ii) le degré de Morley y est définissable. À ma connaissance, et au grand désespoir de Hrushovski, la nécessité de la deuxième condition n'a pas été levée.

J'ai déclaré plus haut que le théorème 5.2 était le résultat principal de l'article ; c'est bien sûr matière à appréciation, influencée par l'évolution ultérieure du sujet. Il faut remarquer que, dans son introduction en russe, Zilber en donne un énoncé différent de celui qui figure dans le corps du texte, en le remplaçant par son corollaire : un groupe algébrique simple infini (sur un corps algébriquement clos) est une structure $\aleph_1$-catégorique.

Dans cette même introduction, il cite en premier son théorème 6.1, seul mentionné dans son résumé en anglais, et qui répond à une question d'Angus Macintyre : *un corps gauche* (тело) *faiblement catégorique est un corps* (поле) *commutatif*. Ce théorème est aussi démontré de façon indépendante dans [Cherlin 1978]. Une fois qu'on sait que le corps gauche L a un sous-corps commutatif infini K, sur lequel L est nécessairement de dimension finie, sa démonstration n'est plus qu'un exercice d'algèbre linéaire. En effet, depuis [Macintyre 1971], on sait que K, comme tout corps infini ω_1-catégorique, et même totalement transcendant, est algébriquement clos. On peut spéculer sur ce qui a manqué à Macintyre pour traiter des corps gauches (finitude du rang, existence d'un sous-corps commutatif infini ?), mais ce qui est sûr, c'est que sa démonstration n'est pas qu'un objet de curiosité pour amateurs d'archéologie mathématique : elle est toujours d'actualité, car pour montrer son théorème on ne connaît aujourd'hui rien d'autre que son appel à la théorie de Galois, ce qui fait qu'on ne sait toujours pas si un corps minimal (*pas* fortement minimal) de caractéristique nulle est algébriquement clos (voir [Wagner 2000a]).

Zilber cite aussi dans l'introduction son théorème 4.2, répondant à une question de Taitslin, déclarant qu'un groupe de théorie universelle $\aleph_1$-catégorique est abélien. Pour cela, il suffit de savoir qu'un groupe infini faiblement catégorique contient un sous-groupe commutatif infini, ce qu'il montre dans son lemme 12, tout en citant la prépublication [Reineke 1975] ; on peut dire de la démonstration de Reineke la même chose que de celle de Macintyre.

On a l'impression que Zilber, mû par un désir de communication, attire l'attention sur les conséquences de ses résultats qui parlent à ses contemporains ; dans la section 3, nous verrons un autre exemple où, pour des raisons de réclame, on met en valeur une conséquence anecdotique en escamotant le résultat profond et novateur, qu'il est plus difficile de leur faire apprécier.

Dans mon exposé, je vais rappeler quelques questions anciennes, mais aussi en introduire de plus personnelles, en prenant le risque qu'elles ne soient ni pertinentes, ni originales. Pour commencer, j'ose surcharger la partition d'un maître en parlant, à la différence de Zilber, d'interprétation de structures et non de théories :

Question 1. (i) *Deux structures ω_1-catégoriques dénombrables sont-elles interprétables dans une même troisième* ?

(ii) *En particulier, deux corps algébriquement clos dénombrables de degrés de transcendance différents sont-ils interprétables dans une même structure ω_1-catégorique* ?

Cette question n'a de sens que pour des structures dénombrables, car une structure ω_1-catégorique ne peut interpréter deux ensembles infinis de cardinaux distincts.

Observons que le successeur des entiers $(\mathbb{Z}, x+1)$ interprète son double $\mathbb{Z}+\mathbb{Z}$ comme sous-ensemble de $\mathbb{Z}\times\mathbb{Z}$, sur la base $(0, \mathbb{Z}) \cup (1, \mathbb{Z})$; en fait, il interprète (avec paramètres) chacune de ses extensions élémentaires non saturées comme un sous-ensemble propre de $\mathbb{Z}\times\mathbb{Z}$, ainsi que son extension dénombrable saturée sur $\mathbb{Z}\times\mathbb{Z}$ tout entier.

Nous examinons maintenant la deuxième apparition de la conjecture, sous la plume de Gregory Cherlin, qui conclut ainsi son article [1979] d'un ton plus assuré que celui de la modeste question (вопрос) de Zilber :

Main Conjecture. *Every simple ω-stable group is an algebraic group over an algebraically closed field.*

Conjecture principale. *Tout groupe simple ω-stable est un groupe algébrique sur un corps algébriquement clos.*

Je crois que si Cherlin n'est pas allé jusqu'à superstable, c'est qu'il n'était pas encore totalement sûr qu'un corps infini superstable fût algébriquement clos (voir la page 2 de son article, et [Cherlin et Shelah 1980]). Il ajoute que le cas de rang de Morley fini lui semble particulièrement important, ainsi que celui des groupes localement finis, et aussi celui des groupes (définissablement) linéaires.

Alors que l'article de Zilber est consacré à des considérations abstraites sur les groupes $\aleph_1$-catégoriques, celui de Cherlin est une étude plus terre à terre des propriétés algébriques des groupes de rang de Morley un, deux et trois ; sa contribution théorique est l'identification des groupes connexes aux groupes de degré de Morley un, par une méthode qui préfigure les arguments de généricité.

Cherlin attire l'attention sur l'existence possible de contre-exemples à sa conjecture de rang de Morley trois, qu'il qualifie de « bad groups » ; disons tout de suite que ces mauvais groupes de rang trois n'ont été éliminés que très récemment par Olivier Frécon [2018].

Cherlin parle dans son introduction de l'article de Zilber, qu'il n'a découvert qu'après avoir achevé le sien ; il exprime son admiration pour le corollaire sur les groupes algébriques simples, qu'il avait obtenu lui-même grâce à une inspection de la structure algébrique de ces groupes, alors que la démonstration de Zilber n'est que pure théorie des modèles. Curieusement, il ne le cite pas à propos de la conjecture.

De nos jours, on a coutume de l'énoncer ainsi, bien qu'aucun de ses deux auteurs ne l'ait formulée exactement en ces termes :

Conjecture d'algébricité [Zilber 1977; Cherlin 1979]. *Un groupe simple de rang de Morley fini est isomorphe à un groupe algébrique sur un corps algébriquement clos.*

Dès son apparition, cette conjecture s'offre à deux regards : est-elle de nature algébrique, structurelle, ou bien de nature modèle-théorique ? Nous avons vu que Cherlin penche vers le premier côté, et Zilber vers le second. Moi-même, dès que je suis entré dans l'arène, j'ai toujours espéré qu'elle serait résolue non pas par une laborieuse classification, mais par une sorte de *general nonsense* modèle-théorique ; j'étais frustré à l'idée que, pour montrer qu'une famille de groupes est algébrique, il fallût nécessairement en donner la liste. La suite des évéènements [1] n'a pas tout à fait répondu à mon attente naïve.

Pour nous exprimer plus clairement, nous faisons la convention suivante. Un groupe de rang de Morley fini sera une structure de groupe *enrichie* : la base de la structure est munie d'une loi de groupe définissable, mais il peut y avoir des ensembles définissables dans cette structure qui ne le sont pas en termes de la seule loi de groupe ; c'est cette structure enrichie qui a un rang de Morley fini. Quand toute la structure est définissable (avec paramètres) à partir de la loi de groupe, nous parlerons de groupe *nu*. Bien que la théorie des modèles ne fasse pas de différence entre les groupes nus et les groupes vêtus, il importe de noter que la conjecture d'algébricité ne parle que du groupe nu.

Et à ce propos, je dois m'excuser d'avoir attribué à Zilber, dans la préface de [Poizat 1987], une conjecture trop enthousiaste, à savoir qu'un groupe simple de rang de Morley fini était un groupe algébrique *nu* sur un corps algébriquement clos, c'est-à-dire qu'il était impossible d'enrichir sa structure en conservant la catégoricité (on sait maintenant que c'est possible grâce à un amalgame de Hrushovski). Cette dernière hypothèse était issue de nos discussions, lorsque j'étais son hôte pendant

1. Écriture inclusive.

un mois à Kemerovo en 1986, mais ne correspond à rien de ce qu'il a écrit ou rendu public.

Cette problématique de nudité renvoie à ce qu'on appelle aujourd'hui la « trichotomie de Zilber », thème d'une autre saga que celle que je chante ici ; voir [Pillay 2013, p. 177; Poizat 2000]. Je ne parlerai pas d'elle, et en particulier pas de son sommet, [Hrushovski et Zilber 1996], où la conjecture est montrée dans un cadre « géométrique » qui reste très abstrait, bien que plus restreint que celui de la finitude du rang de Morley. Je ne parlerai pas non plus d'autres résultats de Zilber, de nature plus algébrique, concernant les groupes de rang de Morley fini, dont plusieurs contributeurs à ce volume rendent compte.

2. Borovik

Un théoricien des groupes étranger à la théorie des modèles, Aleksandr Borovik, dès qu'il a eu connaissance des travaux de Zilber, a voulu dégager une description de son cadre directement accessible à un pur algébriste. Dans [Borovik 1984b], il introduit la liste d'axiomes suivante, définissant ce qu'il appelle les « groupes avec dimension ». Ces conditions ne font intervenir que le groupe lui-même, et pas ses extensions élémentaires.

Аксиома А и аксиома Б. Définition de la collection W des sous-ensembles définissables (конструктивны, *constructibles*, terme emprunté à la géométrie algébrique), avec paramètres, des puissances cartésiennes de G. À chaque constructible A est associé un entier positif $\dim A$.

Аксиома В. $\dim A = 0 \iff A$ — конечно (A est fini).

Аксиома Г. $\dim(A \cup B) = \max\{\dim A, \dim B\}$.

Аксиома Д (принцип связности). Для любого $A \in W$ существует такое число $n \in \mathbb{N}$, что A нельзя представить в виде объединения $n+1$ попарно непересекающихся конструктивных множеств $A_1, \dots, A_{n+1}$ той же размерности, что и A: $\dim A_i = \dim A$.

> **Axiome D** (principe de connexité). Pour chaque $A \in W$ il existe un certain nombre $n \in \mathbb{N}$, tel que A ne puisse pas se présenter sous l'aspect d'une réunion de $n+1$ ensembles constructibles ne se coupant pas par paire $A_1, \dots, A_{n+1}$ de même dimension que A : $\dim A_i = \dim A$.

Аксиома Е (принцип слоев морфизма). Если $A, B \in W$, $f : A \to B$ — морфизм, то множества $B_n = \{x \in B \mid \dim f^{-1}[x] = n\}$ конструктивны и $\dim f^{-1}[B_n] = n + \dim B_n$.

> **Axiome E** (principe à propos des morphismes). Si $A, B \in W$, et $f : A \to B$ est un morphisme, alors l'ensemble $B_n = \{x \in B \mid \dim f^{-1}[x] = n\}$ est constructible et $\dim f^{-1}[B_n] = n + \dim B_n$.

Les axiomes B à Д forcent la dimension de Borovik à majorer le rang de Cantor, mais rien n'impose qu'elle soit le rang de Morley : elle peut être par exemple deux fois le rang de Morley. Ou, plus significativement, si M est interprétable dans une structure $\aleph_1$-catégorique, elle peut être le rang de Morley au sens de la structure-mère et non le rang de Morley intrinsèque de M.

L'axiome E décrit en fait deux propriétés de la dimension : sa définissabilité (pour une famille uniforme de formules avec paramètres) et son additivité. Borovik, citant [Zilber 1974], affirme qu'un groupe G interprétable dans une théorie $\aleph_1$-catégorique satisfait à ces axiomes si on prend pour dimension le rang de Morley au sens de la théorie-mère ; il faut comprendre que G n'est pas muni de sa seule loi de groupe, mais aussi de toute la structure induite, pour garantir la définissabilité du rang. On voit que Borovik, qui n'est pas logicien, et a même l'intention explicite de contourner la théorie des modèles, n'échappe pas à la nécessité d'enrichir les groupes, et il est remarquable que, en ce qui concerne les groupes simples, le résultat de Zilber rende cette précaution inutile.

Une faiblesse de sa présentation, c'est que dans les axiomes A et B, il ne parle que d'images de groupes par des homomorphismes dont les graphes sont constructibles, et jamais de quotient d'un ensemble constructible par une relation d'équivalence constructible. À vrai dire, dans la suite de la prépublication, Borovik parle en une occasion de quotient par un sous-groupe normal *fermé* (замкнутая), un lapsus qui vient de ce qu'il sait qu'un sous-groupe constructible d'un groupe algébrique est Zariski-clos ; il sait aussi qu'un quotient d'un groupe algébrique par un sous-groupe algébrique normal est un groupe algébrique. Nous pensons aller au-devant de ses intentions en qualifiant de *groupe avec dimension* un groupe G habillé dont les ensembles interprétables (avec paramètres) dans G, c'est-à-dire définissables dans la structure G^{eq} obtenue par Shelah en lui ajoutant ses éléments imaginaires, sont munis d'un rang fini satisfaisant aux conditions de Borovik. Cette convention n'a pas d'incidence sur celle de Zilber, pour la raison évidente qu'une structure interprétable dans une théorie $\aleph_1$-catégorique est définissable dans une théorie $\aleph_1$-catégorique obtenue en ajoutant une seule sorte imaginaire à la précédente.

L'additivité du rang de Morley dans un cadre $\aleph_1$-catégorique est bien connue de Zilber, mais il ne l'utilise pas dans [1977], car il n'en a pas besoin pour décrire les propriétés modèle-théoriques des groupes faiblement catégoriques. Cherlin [1979] ne la connaît pas pour les groupes de rang de Morley fini sous sa forme générale ; il se débrouille de façon artisanale, puisque $1+1=2,\ 1+2=3$, et qu'il est inutile de savoir combien font $2+2$ car cela sort du cadre de l'étude.

Voici la version borovikienne de la conjecture d'algébricité :

Гипотеза. Простая бесконечная группа с размерностью является линейной алгебраической группой над алгебраически замкнутым полем.

Conjecture. *Un groupe simple infini avec dimension se trouve être un groupe linéaire algébrique sur un corps algébriquement clos.*

En vrai théoricien des groupes, Borovik n'oublie pas de mentionner qu'un groupe algébrique simple est linéaire. Sa conjecture appelle une triple question :

Question 2. (i) *La structure de groupe nue d'un « groupe avec dimension » au sens de Borovik a-t-elle un rang de Morley fini ?*

(ii) *La conjecture de Borovik est-elle équivalente à la conjecture d'algébricité ?*

(iii) *Qu'en est-il en particulier si le groupe contient un sous-groupe infini d'exposant deux ?*

Les motivations de ces questions s'éclaireront après la section 4, où nous montrerons qu'un groupe, nu ou habillé, de rang de Morley fini satisfait aux conditions de Borovik, la dimension étant le rang de Morley. Si la réponse au premier point est négative, la pertinence du deuxième résidera dans la possibilité de montrer l'équivalence des deux conjectures sans les résoudre ; une réponse négative au troisième permettrait d'évaluer la part de théorie des modèles indispensable à [Altınel et al. 2008] : nous en dirons plus dans la section 6.1.

3. Blum

Les propriétés d'additivité du rang de Morley, dans un cadre général, ont été traitées bien plus tôt dans la première partie du mémoire de doctorat de Lenore Blum [1969] ; j'en extrais les inégalités suivantes dans le cas particulier où les rangs de Morley sont des nombres finis.

(0) Si (a, b) satisfait une formule $\varphi(x, y)$ impliquant $\mathrm{RM}(y/x) \leq m$, alors

$$\mathrm{RM}(a, b) \leq \mathrm{RM}(a) + \mathrm{RM}(b/a) + \mathrm{RM}(a) \cdot m.$$

(1) $$\mathrm{RM}(a, b) \leq \mathrm{RM}(a) + \mathrm{RM}(b/a) + \mathrm{RM}(a) \cdot \mathrm{RM}(b).$$

(2) $$\mathrm{RM}(A \times B) < (1 + \mathrm{RM}(A)) \cdot (1 + \mathrm{RM}(B)).$$

(3) Si tous les 1-types sont de rang de Morley fini, alors tous les n-types sont de rang de Morley fini.

On montre le premier point pour le rang de Cantor par induction sur le rang, puis pour le rang de Morley en montant à un modèle ω-saturé. On en déduit immédiatement les trois dernières inégalités ; elles sont optimales, et ne sont pas valables pour le rang de Cantor, car $\mathrm{RC}(b)$ ne majore pas $\mathrm{RC}(b/a)$. En fait, on peut avoir $\mathrm{RM}(a) = \mathrm{RM}(b/a) = 1$ et $\mathrm{RM}(a, b) = \mathrm{RM}(b) = \mathrm{RM}(M) = \omega$, ou n'importe quoi plus grand ou égal à 2 !

La deuxième partie de la thèse est consacrée aux corps différentiellement clos de caractéristique nulle : c'est là qu'apparaît leur axiomatisation basée sur l'existence de solutions aussi génériques que possible des équations différentielles algébriques en une variable. La question qui gouverne la thèse est celle de la transitivité de la relation « être de rang de Morley fini sur ... », qui n'est pas vérifiée en général, mais qui est valide pour les corps différentiellement clos, où l'ordre de l'équation différentielle minimale borne le rang de Morley.

La partie différentielle de la thèse a été publiée dans [Sacks 1972], mais la première partie, profondément novatrice et plus difficilement appréciable aux contemporains, est restée inédite. C'est la raison pour laquelle [Lachlan 1980] ne la cite pas quand il retrouve ses résultats. Entretemps, Daniel Lascar [1976] avait introduit son rang U et ses fameuses inégalités qui, dans le cas fini, deviennent l'égalité $\mathrm{RU}(a,b) = \mathrm{RU}(a) + \mathrm{RU}(b/a)$.

4. Le petit livre jaune

La préface et le chapitre 2 de [Poizat 1987] apportent au sujet la clarification suivante, en montrant qu'un groupe enrichi G est de rang de Morley fini si et seulement si G^{eq} satisfait aux trois conditions suivantes :

(1) Chaque ensemble définissable a un rang de Cantor fini.

(2) Les cardinaux des membres d'une famille uniforme d'ensembles définissables finis sont bornés (absence de la propriété de recouvrement fini).

(3) Le rang de Cantor est définissable.

Nous insistons sur quelques aspects spécifiques de ces conditions :

(i) Elles ne mentionnent que G lui-même, sans demander de monter à une de ses extensions élémentaires saturées (de fait, il est rare qu'on doive faire appel à des arguments de compacité — le pain quotidien du théoricien des modèles — lors d'études structurelles de groupes de rang de Morley fini).

(ii) L'additivité n'est pas un axiome, c'est une conséquence des axiomes, qui impliquent que le rang de Cantor est aussi le rang de Morley, ainsi que le rang U de Lascar. Elle est toutefois considérée comme un axiome par Borovik et Nesin [1994, p. 57], qui l'utilisent pour montrer des résultats de base, concernant la généricité en particulier, où elle tient lieu de symétrie de la déviation.

(iii) Il en est de même de la stabilité, rarement utilisée en tant que telle (voir [Borovik et Nesin 1994, p. 63]), et de ses conséquences comme la condition de Baldwin–Saxl (que Borovik et Nesin [1994, p. 80] se donnent la peine de montrer à partir de leurs axiomes).

(iv) Si un groupe enrichi les satisfait, il en est de même de n'importe quel groupe interprétable dans la structure, et en particulier du groupe nu associé. Ses extensions élémentaires les satisfont également.

(v) En l'absence de groupe, (1), (2) et (3) n'impliquent que la superstabilité [Burdges et Cherlin 2002] ; à l'inverse, ces propriétés ne sont pas valides dans toutes les structures de rang de Morley fini.

Le petit livre jaune reste jusqu'à ce jour le seul document où est exposée la démonstration de ce fait, qui pourtant n'est ni très compliquée, ni même très originale. Elle repose sur une décomposition du groupe considéré en une tour finie de sous-groupes normaux définissables, chaque quotient étant sous le contrôle d'un ensemble minimal ; elle a été initiée par Lascar [1985] dans le prolongement du théorème de [Zilber 1977] sur les groupes simples.

Son ingrédient principal est le « théorème des indécomposables » de Zilber ([1977, Теорема 3.3; Poizat 1987, p. 44–47] ; voir la section 8), affirmant que, sous certaines circonstances, le sous-groupe engendré par une partie définissable est lui-même définissable, et connexe ; il généralise un résultat, semble-t-il dû à Chevalley, à propos des fermés de Zariski irréductibles d'un groupe algébrique.

Pour appliquer cette décomposition à un groupe infini G de rang de Morley fini, on procède ainsi : on considère, dans une extension élémentaire ω-saturée Γ de G, une partie définissable A fortement minimale ; on l'émonde d'une partie finie pour la rendre indécomposable, puis on la translate pour qu'elle contienne l'élément neutre, et engendre un groupe définissable connexe. Les conjugués de ce groupe engendrent un groupe définissable connexe normal Γ_1, et on itère le procédé dans Γ/Γ_1, qui est de rang de Morley inférieur (on n'a pas besoin de l'additivité pour le montrer !) ; on est arrêté au bout d'un nombre fini de pas, pour constater qu'on a affaire à une structure fini-dimensionnelle dont chaque dimension est portée par un ensemble fortement minimal. Dans une telle structure, le rang de Cantor est égal au rang de Morley, au rang U et au poids, et la condition d'uniformité est vérifiée, comme c'est le cas pour une structure $\aleph_1$-catégorique, c'est-à-dire dans le cas unidimensionnel [Baldwin 1973; Belegradek 1973; Poizat 1978] ; dans ces structures, le degré de Morley d'une famille uniforme d'ensembles définissables est borné, mais pas nécessairement définissable [Hrushovski 1992]. Cerise sur le gâteau : il s'agit de propriétés de la *théorie* de Γ, qui est aussi celle de G, si bien que la décomposition de Lascar peut se conduire dans G lui-même, qu'il est a posteriori inutile de le remplacer par un modèle saturé.

Réciproquement, si un groupe satisfait aux trois conditions, ses ensembles définissables infinis minimaux le sont fortement grâce à la condition 2 d'uniformité, c'est-à-dire qu'ils restent minimaux dans ses extensions élémentaires. On conduit alors la décomposition de Lascar en s'appuyant sur le rang de Cantor pour montrer la fini-dimensionnalité.

On voit de même que, si un groupe avec dimension au sens de Borovik satisfait à la condition 2 d'uniformité, par exemple s'il est ω-saturé, la décomposition de Lascar (reposant sur la dimension) montre que ce groupe est de rang de Morley fini.

Dans ce cas, il est clair qu'on obtient une fonction de dimension sur chaque extension élémentaire de ce groupe G en faisant usage de la définition de la dimension au sens de G.

Ce qui a été dit plus haut du degré de Morley rend nécessaire de rappeler une double question qui ne m'appartient pas, liée à l'obstacle qu'a rencontré Hrushovski dans ses amalgames [1992].

Question 3 [Borovik et Nesin 1994, p. 371]. *Dans un groupe de rang de Morley fini :*

(i) *Le degré de Morley est-il définissable ?*

(ii) *Les composantes connexes d'une famille uniforme de groupes définissables forment-elles une famille uniforme ?*

5. Berline et Lascar

Chantal Berline et Daniel Lascar [Berline 1986; Berline et Lascar 1986] montrent que, dans l'étude d'un groupe superstable G, le développement de Cantor de son rang, $\mathrm{RU}(G) = \omega^{\alpha k} \cdot n_k + \cdots + \omega \cdot n_1 + n_0$, joue un rôle essentiel. Tout d'abord, il y a des types de RU maximal, qui sont les types génériques, si bien que l'écriture $\mathrm{RU}(G)$ a un sens ; ensuite, grâce à une version infinitaire du théorème des indécomposables, ils obtiennent une décomposition de Lascar de G en une tour $G \supset G_k \supset \cdots \supset G_1 \supset G_0$ de groupes définissables normaux dont les quotients consécutifs G_{i+1}/G_i ont un RU monomial. C'est ainsi qu'un groupe superstable simple doit avoir un RU monomial, $\mathrm{RU}(G) = \omega^\alpha \cdot n$, et doit contenir un sous-groupe abélien de rang au moins ω^α, et qu'un corps gauche superstable est commutatif. Tout ceci est aussi exposé dans [Poizat 1987, chapitre 6].

Ils reprennent ensuite les résultats de Cherlin pour les groupes de rang ω^α, $\omega^\alpha \cdot 2$ et $\omega^\alpha \cdot 3$. Maintenant qu'on sait, grâce à Frécon, que les « bad groups » de rang 3 n'existent pas, il est naturel de poser la question de l'extension de son résultat au cas superstable.

Question 4. *Peut-on adapter la démonstration de Frécon aux groupes superstables de rang $\omega^\alpha \cdot 3$?*

Comme préalable à la question, il faudra décider de quel rang on parle. L'unicité de la racine carrée dans un groupe de rang de Morley fini sans involutions jouant un tel rôle dans la démonstration de Frécon, il faudra vraisemblablement se placer dans un contexte ω-stable.

6. Wagner and the ABC murders

6.1. ***Imiter la classification des groupes simples finis.*** L'idée-force de ce qu'on appelle maintenant le « programme de Borovik » est d'attaquer la conjecture en adaptant les techniques employées lors de la classification des groupes simples finis, en s'appuyant principalement sur le comportement des involutions dans un contexte de rang de Morley fini. Borovik, dès [1984a], entreprend de généraliser des propriétés bien connues des involutions d'un groupe fini, mais le premier résultat de quelque ampleur est la conjugaison des 2-sylows, montrée dans [Borovik et Poizat 1990] ; [Poizat et Wagner 1993] en donne une démonstration plus algébrique. Comme en 1987 je mettais en doute la possibilité de montrer quelque chose à propos de sous-groupes non définissables, mon coauteur m'a répondu que ce qui importait n'était pas qu'ils fussent définissables ou pas, mais que c'étaient les 2-sylows.

Cette approche algébrique a ensuite donné lieu à une somme considérable de travaux, dus à une multitude d'auteurs, dont je ne rends pas compte ici (pas plus que de leurs généralisations à des contextes plus larges, comme le cadre o-minimal) ; une première génération d'entre eux est exposée dans [Borovik et Nesin 1994].

Elle culmine dans [Altınel et al. 2008], où la conjecture est montrée pour les groupes dont les 2-sylows sont non triviaux et ont un exposant borné. Ce monumental ouvrage donne aussi des informations sur les groupes contenant des 2-groupes de Prüfer, mais dit peu de choses sur les possibles groupes simples de rang de Morley fini sans involutions (incompatibles avec la conjecture), qu'il réussit à contourner.

Ses résultats sont principalement de nature algébrique, structurelle, mais ils sont conditionnés par un théorème d'essence modèle-théorique, une conséquence très inattendue des axiomes caractérisant le rang, le théorème de [Wagner 2001] sur les corps de rang de Morley fini (K algébriquement clos, avec structure supplémentaire), déclarant que : (i) K élimine les imaginaires ; (ii) son modèle premier est la clôture algébrique modèle-théorique de $\varnothing$; (iii) si K a un automorphisme définissable non trivial, ce modèle premier est porté par la clôture algébrique algébrique de $\varnothing$. Il a pour conséquence qu'en caractéristique p un tore (c'est-à-dire un sous-groupe définissable de $K^* \times \cdots \times K^*$) est clôture définissable de sa torsion.

6.2. ***Utiliser la classification des groupes simples finis.*** La classification des groupes simples finis a par elle-même des conséquences sur la conjecture d'algébricité : c'est en s'appuyant sur elle que Simon Thomas [1983] a montré que cette conjecture est vraie pour les groupes localement finis. Il a ensuite observé qu'elle était aussi vraie pour les groupes pseudo-localement finis.

Une structure est *localement finie* si chaque partie finie de sa base engendre une sous-structure finie, et *pseudo-localement finie* si c'est un modèle de la théorie des structures localement finies de même langage.

Cette définition repose sur une convention linguistique, puisque toute structure de langage purement relationnel est localement finie. Pour l'éviter, nous introduisons le solide *principe de finitude locale*, qui, par définition, est satisfait par les structures dans lesquelles toute structure interprétable (avec paramètres) est pseudo-localement finie. Voici comment ce principe s'introduit dans notre sujet :

Fait. *Les corps algébriquement clos nus*, *ainsi que les corps de Wagner* (*avec automorphisme*) *mentionnés ci-dessus*, *satisfont au principe de finitude locale.*

Les ingrédients de la démonstration sont l'élimination des imaginaires, et le fait que les corps finis sont définissablement clos (si un détail vous échappe, consultez [Poizat 2001b]).

Après avoir constaté que les groupes définissablement linéaires en caractéristique positive sont définissables dans un corps de Wagner, [Poizat 2001b] en conclut que :

Corollaire. *La conjecture d'algébricité est vraie pour les groupes définissablement linéaires en caractéristique p.*

Cela mène inévitablement à la question suivante, posée aussi dans [Macpherson et Pillay 1995] :

Question 5. *Est-elle vraie pour un groupe définissablement linéaire en caractéristique nulle* ? (*Il faut éliminer un mauvais groupe qui ne contient que des éléments semi-simples* ; *ce serait un groupe linéaire simple sans involutions*, *dont aucun exemple n'est connu à ce jour.*)

Le principe de finitude locale implique le *principe de surjectivité*, dit aussi *principe d'Ax* [1968], déclarant que toute injection définissable d'un ensemble définissable dans lui-même est surjective. Cela nous rappelle une question ancienne qui n'a de sens que pour les groupes nus, car on peut enrichir un corps en lui amalgamant une extension élémentaire du successeur des entiers naturels :

Question 6 [Borovik et Nesin 1994, p. 371]. *Un groupe nu de rang de Morley fini satisfait-il au principe de surjectivité* ?

Il est bien des contextes où la pseudo-finitude locale permet d'étendre immédiatement, ou presque, aux groupes de rang de Morley fini des propriétés des groupes finis, par exemple celui des groupes de Frobenius [Poizat 2024]. Nous pouvons rêver que c'est dans cette pseudo-finitude locale que nous trouverons le chaînon manquant entre ces deux familles de groupes, et croire que postuler la conjecture d'algébricité revient à dire que tout groupe de rang de Morley fini satisfait au principe de finitude locale ; cela nous donnerait une formulation de la conjecture qui s'abstiendrait de privilégier les groupes simples. Avant de planer dans ces hauteurs, il est prudent de réfléchir à une question plus terre à terre :

Question 7. *En caractéristique nulle*, *les corps verts de rang de Morley fini*, *et plus précisément ceux dont l'existence est assurée par* [Baudisch et al. 2009], *sont-ils pseudo-localement finis ?*

Rappelons qu'on ne sait pas s'il y a des « corps de Wagner » autres que les corps nus et que, d'après [Wagner 2003], l'existence de corps verts en caractéristique p aurait des conséquences arithmétiques surprenantes sur les nombres de Mersenne.

7. Weil et Hrushovski

Le travail de Thomas consiste en définitive, grâce au principe de finitude locale, à déduire la classification des groupes algébriques simples de celle des groupes simples finis, ce qui ne manque pas de paradoxe ; il a pour conséquence que la conjecture n'a pas de contre-exemple parmi les groupes définissables dans un corps algébriquement clos nu !

C'était sans doute la première chose à vérifier dès qu'elle est apparue, mais personne n'a songé à le faire à ce moment. De fait, il n'y a pas besoin de mobiliser la classification des groupes simples finis pour cela, car tout groupe définissable dans un corps algébriquement clos est définissablement isomorphe à un groupe algébrique ; ce n'est pas tout à fait évident, et demande l'adaptation par Hrushovski d'un théorème de [Weil 1955].

En effet, un groupe définissable connexe permet de définir génériquement des morphismes représentant la restriction de la multiplication et celle de l'inverse. C'est évident en caractéristique nulle, et demande un petit travail en caractéristique p ; comme l'a remarqué Laurentius van den Dries, on peut alors faire appel à un théorème d'André Weil, déclarant qu'un « groupe partiel », un « group chunk », correspond à une partie générique d'un groupe algébrique [van den Dries 1982]. On trouvera dans [Poizat 1987, chapitre 4(e)] un exposé du résultat accessible à un théoricien des modèles, qui doit beaucoup à Hrushovski (voir aussi les notes historiques aux pages 154–155).

Le chapitre 5(f) de [Poizat 1987] expose ce qu'on peut considérer comme la version modèle-théorique ultime du théorème de Weil, démontrée par [Hrushovski 1986] dans un contexte stable : une opération binaire qui satisfait associativité générique et simplificabilité générique (à gauche et à droite) correspond à une partie générique d'un groupe définissable.

Le chapitre 2(f) décrit une décomposition d'un groupe de rang de Morley fini en une tour de groupes définissables, basée sur la notion d'internité au sens de Hrushovski, qui, à la différence de la décomposition de Lascar, se fait à partir du haut. Sa première étape consiste à choisir un type q régulier, non orthogonal aux types génériques du groupe G, ce qui permet d'obtenir un sous-groupe G_1 définissable et normal dans G tel que le quotient G/G_1 soit q-interne.

Le chapitre 4(g) donne une version modèle-théorique du théorème de Borel–Tits basé sur l'internité : si G est un groupe algébrique simple sur un corps algébriquement clos K, on peut définir dans G un corps K_1 tel que G soit K_1-interne. À partir du seul résultat de nature algébrique affirmant que, puisque K_1 est définissable dans le corps nu K, il lui est nécessairement définissablement isomorphe, on en déduit que tout ensemble de G^{eq} qui est définissable au sens de K est en fait définissable au sens du groupe nu G.

C'est peut-être un bon endroit pour rappeler un cas particulier où, à ce qu'il me semble, la conjecture d'algébricité est toujours ouverte :

Question 8 [Borovik et Nesin 1994, p. 367; Kramer et al. 1999; Mustafin et Poizat 2006]. *Soient $G(K)$ un groupe algébrique sur un corps algébriquement clos K, et L un sous-corps infini de K contenant le corps de définition de $G(\cdot)$. On note $G(L)$ le groupe des points L-rationnels de $G(\cdot)$. Si $G(L)$ est simple et de rang de Morley fini, est-ce que L est algébriquement clos ?*

L'introduction de [Kramer et al. 1999] explique pourquoi c'est vrai quand le corps L est localement fini.

8. Génération elliptique et théorème sans indécomposables

Nous dirons qu'une partie A du groupe G engendre *elliptiquement* le sous-groupe H s'il existe un entier m tel que chaque point de H soit le produit de m points de $A \cup \{1\} \cup A^{-1}$ (cette définition ne suppose pas que H soit tout le groupe engendré par A).

Le « théorème des indécomposables » de [Zilber 1977] affirme justement que le sous-groupe engendré par l'ensemble définissable A, indécomposable et contenant l'élément neutre, l'est de façon elliptique, ce qui le rend définissable (il est de plus connexe). Il fait intervenir une notion *ad hoc* d'ensemble indécomposable, généralisant la notion de fermé de Zariski irréductible en géométrie ; dans bien des situations concrètes, son usage demande de l'ingéniosité, car on doit utiliser des propriétés de l'ensemble A pour restreindre la famille des groupes pour lesquels il suffit de tester l'indécomposabilité. Il est plus pratique de le remplacer par l'énoncé équivalent suivant, qui est plus flexible justement parce qu'il ne mentionne pas l'indécomposabilité, ni aucune autre propriété des ensembles définissables (on le montre facilement à partir du théorème des indécomposables, comme dans [Poizat 2021]. Sa version première apparaît dans [Wagner 2000b, Chapter 4] dans un contexte plus général).

Théorème sans indécomposables. *On considère une famille $\{A_i\}$ de parties définissables d'un groupe G de rang de Morley fini ; il existe alors un plus grand sous-groupe H de G qui soit définissable, connexe, et elliptiquement engendré par la réunion d'un nombre fini de A_i. De plus, chaque A_i normalise H et se répartit en un nombre fini de classes modulo H.*

Je ne résiste pas au plaisir d'en illustrer la facilité d'emploi par un petit corollaire :

Corollaire. *Le groupe engendré par un ensemble A définissable clos par conjugaison et par élévation à la puissance m^e, pour un $m > 1$, est définissable, car elliptiquement engendré.*

Démonstration. On montre que tout produit d'éléments de A peut se remplacer par un produit où chaque point ne se répète que moins de m fois ; on en déduit que A engendre un groupe fini modulo le groupe H ci-dessus. □

Exemples d'application. Les puissances n^{es}, les solutions de l'équation $x^n = 1$, les produits de deux involutions sont clos par prise de carré. Les éléments d'ordre exactement n sont clos par puissance p^{e}, pour chaque p premier à n. Si S est un ensemble convexe (c'est-à-dire clos par symétries ; voir la section 9) contenant l'élément neutre, $S \cdot S$ est clos par conjugaison et prise de carré (voir [Poizat 2018, proposition 13]). Le groupe dérivé G' du groupe G de rang de Morley fini est elliptiquement engendré par les commutateurs, car $G' \times \{1\}$ est l'intersection avec $G \times \{1\}$ du sous-groupe de $G \times G$ engendré par le « convexe symétrique » formé des (x, x^{-1}) [Poizat 2018].

La question se pose d'une sorte de réciproque, à savoir : *si une partie définissable engendre un sous-groupe définissable*, *est-ce-que cette génération est elliptique* ? Modulo le théorème sans indécomposables, une question équivalente est :

Question 9 [Deloro 2007; Ould Houcine 2007; Poizat 2021]. *Existe-t-il un groupe infini*, *de rang de Morley fini*, *qui soit finiment engendré* ?

Son étrangeté est de mettre en scène des groupes de rang de Morley fini non saturés, alors que Borovik comme le petit livre jaune nous ont habitués à vivre sans le souci des propriétés de saturation du seul modèle que nous avons à l'horizon.

Une réponse positive contredirait la conjecture d'algébricité, car :

• Si G est finiment engendré, G° l'est également, puisqu'il est d'indice fini dans G.

• Si on divise G° par un groupe normal définissable connexe maximal M, on ne peut pas obtenir un groupe commutatif : en effet, comme il est finiment engendré, ce serait le produit d'un nombre fini de groupes cycliques, et comme il est infini il ne serait divisible par aucun nombre premier, alors qu'il devrait l'être par presque tous.

• G/M est donc, à un centre fini près, un groupe simple ; mais un groupe simple infini et finiment engendré ne peut être linéaire, car selon [Maltsev 1940] un groupe linéaire finiment engendré est résiduellement fini (c'est-à-dire que l'intersection de ses sous-groupes d'indice fini est réduite à l'élément neutre).

9. D'autres questions désespérées

La conjecture d'algébricité, si elle se confirme, aurait un certain nombre de conséquences de nature purement algébrique sur la structure nue d'un groupe quelconque de rang de Morley fini. Pour s'en rendre compte, la règle du jeu est de considérer un contre-exemple minimal, et de montrer qu'il s'agit d'un groupe simple, qui serait linéaire d'après la conjecture, et d'utiliser des propriétés bien connues en pure théorie des groupes, comme :

— Tout groupe linéaire finiment engendré est résiduellement fini.
— Tout groupe linéaire périodique est localement fini.
— Tout groupe infini, localement fini, et satisfaisant à la condition de chaîne sur les centralisateurs, contient un sous-groupe abélien infini.

On introduit alors ces propriétés sous forme de questions, auxquelles je prédis qu'il y a peu d'espoir d'apporter une réponse sans résoudre la conjecture :

Question 10. *Un sous-groupe périodique d'un groupe de rang de Morley fini est-il localement fini ?*

Question 11. *Un sous-groupe infini d'un groupe de rang de Morley fini contient-il un groupe abélien infini ?*

Question 12 [Borovik et Nesin 1994, p. 368]. *Un groupe de rang de Morley fini est-il localement résiduellement fini ?*

10. Zamour

La section finale de cet article élargit la problématique de la conjecture d'algébricité à des structures plus faibles que des groupes. Même si l'on pense que les résultats dont je vais parler ici ont un côté anecdotique, il est du moins certain qu'ils éclairent la nature modèle-théorique si particulière des groupes de rang de Morley fini ; les questions posées seront plus naïves, car, à la différence des précédentes, ce ne sont pas des problèmes qui ont cassé toutes les têtes depuis cinquante ans.

Dans un groupe G, l'opération binaire $s(x, y) = yx^{-1}y$ satisfait aux trois équations suivantes : (i) $s(x, x) = x$; (ii) $s(s(x, y), y) = x$; (iii) $s(s(x, z), s(y, z)) = s(s(x, y), z)$. Si on convient d'appeler, une fois y fixé, l'opération unaire $s_y(x) = s(x, y)$ *symétrie de centre* y, ces équations ont l'interprétation transparente suivante : (i) chaque symétrie fixe son centre (ou plus exactement fixe chacun de ses centres) ; (ii) chaque symétrie est une bijection involutive ; (iii) chaque symétrie est un automorphisme de la structure.

Une loi binaire satisfaisant à ces trois conditions est appelée *espace de symétries* ; dans [Poizat 2021], j'ai rebaptisé *symétron* ceux de ces espaces dans lesquels, pour tous x et y, il existe un unique z tel que $s(x, z) = y$; il est appelé le *milieu* $m(x, y)$ de x et de y.

L'espace des symétries du groupe G est un symétron si et seulement si ce dernier est uniquement 2-divisible, c'est-à-dire si chaque x possède une unique racine carrée $x^{1/2}$. C'est le cas de tous les groupes de rang de Morley fini, ou plus généralement de tous les groupes ω-stables, qui n'ont pas d'involutions (car alors le centre du centralisateur de x est 2-divisible). Dans le contexte du rang de Morley fini, ces symétrons sont apparus dans [Poizat 2018] comme un outil exégétique pour commenter la démonstration de [Frécon 2018] ; ils interviennent aussi dans la structure des groupes de Frobenius ayant une involution dans leur complément [Clausen et Tent 2023; Poizat 2024].

Dans [Poizat 2018], on étudie les sous-symétrons définissables d'un groupe de rang de Morley fini. Par contre, dans [Poizat 2021], les symétrons de rang de Morley fini sont considérés comme des structures *per se*. Il est constaté que bien des propriétés qu'on montre habituellement pour les groupes se généralisent à ces symétrons : condition de chaîne (car, pour une partie définissable d'un symétron ω-stable, il est équivalent d'être clos par symétrie ou d'être clos par prise de milieu), décomposition en composantes connexes, ainsi que des théorèmes de génération elliptique. Toute cette théorie repose sur une base fragile (faut-il évoquer le songe que Daniel a dû interpréter ?), à savoir que, dans un symétron ω-stable, deux points sont toujours reliés par un sous-symétron isomorphe à celui des symétries d'un groupe commutatif définissable sans involutions. On en déduit qu'un sous-symétron définissable est le symétriseur de l'ensemble de ses types génériques (de même qu'un sous-groupe définissable est le stabilisateur, par translations à droite comme par translations à gauche, de l'ensemble de ses types génériques ; cette description des sous-symétrons définissables par des données génériques permet de court-circuiter les calculs de Frécon).

Mais que se passe-t-il si on considère des lois valables génériquement, c'est-à-dire deux opérations partielles $s(x, y)$ et $m(x, y)$ qui satisfont génériquement les équations de symétrons, sans qu'on soit déjà enveloppé à l'intérieur d'un symétron ? Autrement dit :

Question 13. *Existe-t-il un théorème de Weil–Hrushovski pour les symétrons ?*

L'acuité de la question est renforcée par le fait qu'il existe des structures assez voisines pour lesquelles le théorème de Weil est faux [Grishkov et Nagy 2011]. La question suivante lui est probablement liée :

Question 14. *En caractéristique autre que* 2, *un symétron constructible est-il définissablement isomorphe à un symétron algébrique*, *basé sur une variété algébrique pour laquelle symétrie et milieu sont des morphismes ?*

Ce n'est pas vrai en caractéristique 2, car dans le groupe multiplicatif K^* la racine carrée de x, qui n'est pas un morphisme, est le milieu de 1 et de x.

[Borovik et Nesin 1994, p. 355] demandent si le théorème Z^* de Glauberman — un classique de la théorie des groupes finis — s'étend aux groupes de rang de Morley fini. Cela se traduit par la question suivante, à laquelle le théorème 1 de [Glauberman 1966] apporte une réponse positive dans le cas des symétrons finis.

Question 15. *Dans un symétron de rang de Morley fini*, *est-ce que le dérivé du groupe de permutations engendré par les symétries est dépourvu d'involutions ?*

Cette question, qui ne concerne que le symétron nu, a un lien étroit avec la conjecture d'algébricité. En effet, d'après [Poizat 2018; 2021], d'une part, la réponse est positive pour les symétrons pseudo-localement finis, et d'autre part, quand le groupe engendré par les symétries est définissable (on parle alors de symétron *borné*), ce qui est le cadre de la question de Borovik et Nesin, une réponse négative contredit la conjecture d'algébricité.

Venons-en finalement à la question que s'est posée Samuel Zamour dans la première partie de sa thèse, dont je n'ai eu connaissance que quelques jours avant sa soutenance. Elle est purement théorique, car d'après la proposition 13 de [Poizat 2018], un sous-symétron définissable d'un groupe de rang de Morley fini est inter-définissable avec un groupe habillé, et il semble peu probable que les symétrons non bornés interviennent un jour dans la solution de la conjecture d'algébricité (pour les groupes sans involutions).

Question 16 [Zamour 2022]. *Les symétrons rangés*, *c'est-à-dire satisfaisant aux conditions* 1, 2 *et* 3 *de la section 4*, *sont-ils les mêmes que les symétrons de rang de Morley fini* ?

Zamour reprend la problématique de Zilber à son début, car sa question est très proche du théorème 5.2 de [Zilber 1974] ; on s'en rend compte si on substitue au mot группа le mot симметрон dans son énoncé :

Теорема 5.2. Пусты G — простая слабо категоричная группа. Тогда теория G в групповой сигнатуре $\aleph_1$-категорична, т.е. G — категоричная группа.

> **Théorème 5.2.** Soit G un groupe simple faiblement catégorique. Alors la théorie de G dans le langage des groupes est $\aleph_1$-catégorique, c'est-à-dire G est un groupe catégorique.

C'est aussi vrai de tout ce qui l'accompagne, comme le théorème 5.1 : un symétron faiblement catégorique définissablement simple (sans congruence non triviale définissable ; voir ci-après) est-il simple ?

Zamour cherche à reproduire une décomposition de Lascar pour apporter une réponse positive, mais il ne réussit que partiellement pour les raisons que nous allons exposer. Il montre d'abord, pour les symétrons, une version du théorème des indécomposables sous sa forme classique : il déclare qu'un ensemble définissable est *indécomposable* si, chaque fois qu'il est contenu dans une réunion finie de sous-symétrons définissables deux à deux disjoints, il est inclus dans l'un d'entre eux ; il montre que le sous-symétron engendré (par symétrie et prise de milieu) par un ensemble indécomposable est définissable, connexe et elliptiquement engendré.

Cela lui permet de franchir sans peine la première étape de la décomposition, c'est-à-dire la construction d'un sous-symétron connexe elliptiquement engendré par un ensemble fortement minimal ; mais il rencontre un obstacle dès qu'il s'agit de faire un quotient. Appelons *congruence* de symétrons une relation d'équivalence définissable telle que la symétrie passe au quotient : si $x \sim x'$ et $y \sim y'$, alors $s(x, y) \sim s(x', y')$. Les classes d'équivalence d'une congruence sont des sous-symétrons qui se correspondent par symétries : chacune d'entre elles détermine la relation $\sim$; nous les qualifions de sous-symétrons *normaux*. Comme la conjonction de deux congruences est une congruence, la condition de chaîne montre que chaque sous-symétron a une clôture normale définissable, qui est le plus petit sous-symétron définissable et normal le contenant.

Dans le cas où les applications $m(x, a)$ sont des automorphismes, ce qui se produit rarement (par exemple dans les systèmes de triplets de Steiner que sont les symétrons où milieu et symétrie coïncident), Zamour montre que la clôture normale d'un sous-symétron S définissable et connexe est elliptiquement engendrée par S et par un nombre fini de paramètres. Cela lui permet de conduire la décomposition de Lascar comme dans le cas des groupes.

À vrai dire, afin de bénéficier du savoir accumulé par les théoriciens des groupes dans l'étude des symétries des groupes finis, Zamour travaille avec des Z-boucles uniquement 2-divisibles, qui sont des structures connues pour être bi-interprétables avec les symétrons ; par exemple, dans le cas des groupes, la boucle est l'opération $y^{1/2} \cdot x \cdot y^{1/2}$. Ces boucles sont mieux aimées de bien des théoriciens des groupes, qui chérissent les structures qui ressemblent le plus possible à des groupes, aussi bizarres soient-elles ; pour ma part, je préfère le langage des symétrons (symétrie et milieu), qui donne une axiomatique claire, et permet d'exprimer naturellement les propriétés spécifiques aux symétrons de rang de Morley fini.

En vertu du principe dangereux qu'il est permis de conjecturer ce qu'on ne sait pas démontrer, on peut poser une dernière question :

Question 17. *Si S est un sous-symétron définissable connexe d'un symétron de rang de Morley fini, est-ce que sa clôture normale est elliptiquement engendrée par S et par un nombre fini de paramètres ?*

Afin d'éviter d'attirer sur nos têtes le courroux du Grand Roi, nous allons prudemment conclure l'article par trois théorèmes ; le premier est la version symétrique du théorème sans indécomposables, et le deuxième une élucidation de l'obstacle qu'a rencontré Zamour dans son étude de la question ci-dessus.

Si A est une partie du symétron S, nous appellerons *composantes* du sous-symétron Σ engendré par A les sous-symétrons définissables connexes maximaux qui sont elliptiquement engendrés par A (en usant du milieu et de la symétrie) ; dans le cas des groupes, leurs analogues seraient les cossettes modulo le plus grand sous-groupe définissable connexe elliptiquement engendré par A.

Théorème. *Soient S un symétron de rang de Morley fini et A un sous-ensemble de S définissable. Alors, le sous-symétron engendré par A est réunion de ses* *composantes ; celles-ci sont deux à deux disjointes, se correspondent par symétries, et A est contenu dans la réunion d'un nombre fini d'entre elles.*

Démonstration. C'est évident quand A est vide. Sinon, chaque point de Σ constitue un symétron définissable connexe elliptiquement engendré par A, et s'étend en un tel sous-symétron de rang de Morley maximal. Comme, d'après [Poizat 2021], deux sous-symétrons définissables connexes d'intersection non vide en engendrent un troisième, et ce de façon elliptique, deux composantes non disjointes sont égales. Si a est un point de la composante C et a' un point de la composante C', la symétrie ayant pour centre le milieu m de a et de a' échange C et C'. Le dernier point vient de ce que, d'après [Zamour 2022], A s'écrit comme réunion d'un nombre fini d'ensembles définissables indécomposables. □

Nous dirons que la relation d'équivalence définissable $\sim$ est une *demi-congruence* si chaque symétrie en est un automorphisme, c'est-à-dire si $x \sim x'$ implique $s(x, y) \sim s(x', y)$. Ses classes seront qualifiées de sous-symétrons *semi-normaux.*

Lemme. (i) *Une demi-congruence est une congruence si et seulement si le milieu passe au quotient, c'est-à-dire si $x \sim x'$ implique que $m(x, y) \sim m(x', y)$.*

(ii) *Le sous-symétron définissable non vide S est semi-normal si et seulement si, quels que soient les points a, b et c, $s_a\big(s_b(s_c(S))\big)$ est égal à S ou disjoint de S.*

Démonstration. (i) Dans une demi-congruence, toutes les classes ont même rang et même degré de Morley. Si c'est aussi une congruence pour le milieu, quel que soit le point a et la classe C, $m(C, a)$ est inclus dans une classe C' ; si b est un point de C', $m(C'', a)$ est aussi inclus dans $m(C, a)$, où C'' désigne la classe de b : comme $m(x, a)$ est une bijection, il faut que $C = C''$; autrement dit $m(x, a)$ définit une bijection entre C et C'. Le même raisonnement vaut pour la bijection inverse $s(a, x)$.

(ii) La propriété est possédée par un sous-symétron S semi-normal, car alors $s_a\big(s_b(s_c(S))\big)$ et S sont des classes de la relation d'équivalence $\sim$ associée.

Réciproquement, on considère deux points a et b, un point x dans S et le milieu m de x et de $s_a(s_b(x))$; comme il ne sont pas disjoints, S et $s_m\big(s_a(s_b(S))\big)$ sont égaux, soit encore $s_m(S) = s_a(s_b(S))$.

Considérons maintenant un translaté $S' = t(S)$ de S par un produit de symétries t : $s_a(s_b(S')) = s_a \cdot s_b \cdot t \cdot S = t \cdot t^{-1} s_a t \cdot t^{-1} s_b t \cdot S = t \cdot s_c \cdot S = t s_c t^{-1}(S')$, où c est le milieu de $t(a)$ et de $t(b)$. Autrement dit, si s' et s'' sont deux symétries, il en existe une troisième s telle que $s(S') = s's''(S')$.

On en déduit que chaque translaté $t(S)$ est de la forme $s(S)$, et que les translatés de S ont la même propriété. Or S et $s(S)$ sont égaux si le centre de s est dans S, et disjoints s'il n'y est pas, puisque S est clos par prise de milieu. En résumé, les translatés de S forment bien une partition. $\square$

Théorème. *Soient S un sous-symétron définissable et connexe d'un symétron de rang de Morley fini, et Σ le plus grand sous-symétron définissable et connexe contenant S qui soit elliptiquement engendré par S et un nombre fini de paramètres. Alors Σ est semi-normal, et pour tout a et toute classe C de la demi-congruence $\sim$ associée, il existe une (unique) classe C', telles que $m(x, a)$ définisse une bijection entre une partie générique de C et une partie générique de C'. De même il existe une unique classe C'' telle que $s(a, y)$ définisse une bijection entre une partie générique de C et une partie générique de C''. Ce résultat est aussi valable pour un symétron rangé.*

Démonstration. Comme deux sous-symétrons connexes d'intersection non vide s'amalgament de façon elliptique, Σ est bien unique, et contient tout sous-symétron définissable connexe et elliptiquement engendré par S et un nombre fini de paramètres qui le coupe ; il satisfait donc au critère du lemme et définit une demi-congruence $\sim$.

Considérons la composante Σ' du symétron engendré par $m(\Sigma, a)$ qui intersecte génériquement l'ensemble $m(\Sigma, a)$, et le milieu m d'un point de Σ' et de Σ ; Σ' est connexe et elliptiquement engendré par S et un nombre fini de paramètres, si bien que $s(\Sigma', m)$ est inclus dans Σ. Il lui est en fait égal, car son rang majore celui de Σ ; Σ' est donc une classe de la demi-congruence $\sim$, et la bijection $m(x, a)$ échange des parties génériques des deux classes Σ' et Σ. Le même raisonnement vaut pour chaque symétrisé de Σ.

Pour le dernier point, il faut vérifier que tout ce qui est utilisé pour la démonstration du théorème, dans cet article et dans [Poizat 2021], est aussi valable dans le cas rangé, ce qui n'est qu'affaire de patience. $\square$

Mea culpa, mea culpa, mea maxima culpa ! Je profite de l'occasion qui m'est offerte (Спасибо, Борис!) pour confesser deux péchés commis dans [Poizat 2021]. Le premier est que j'y affirme gratuitement l'énoncé du théorème suivant :

Théorème. *La partition en composantes connexes d'un symétron ω-stable est une congruence.*

Démonstration. Elle est évidemment une demi-congruence. Pour chaque a du symétron S, la symétrie de centre a permute les composantes du symétron, si bien qu'à toute permutation π est associé l'ensemble S_π des points de S qui permute ses composantes selon π. Les S_π non vides constituent une partition de S ; si les symétries s_a et s_b échangent les ensembles A et B, il en est de même de $s_b s_a s_b$, ce qui signifie que les points dont la symétrie associée échange A et B constituent un sous-symétron ; par conséquent les S_π sont des symétrons définissables. Comme les composantes sont closes par prise de milieu, un point a fixe sa propre composante et seulement celle-là. Par conséquent, si S_π n'est pas vide, π fixe un point et un seul ; donc, si S_π a même rang de Morley que S, il ne contient qu'un seul type générique, et c'est une composante connexe de S. L'unicité de la décomposition de S en un nombre fini de sous-symétrons définissables connexes deux à deux disjoints impose aux S_π d'être ces composantes connexes. Autrement dit, tous les points d'une même composante agissent sur les autres de la même façon. □

Le deuxième est que j'ai omis de préciser, à la fin de l'exemple 2, que a devait être central dans G. On obtient donc un exemple avec un centre non trivial en prenant pour G un groupe de rang de Morley fini, nilpotent non abélien, et sans involutions.

Remerciements

Орысша түйінісі үшін Ержанға үлкен рақмет! Je tiens aussi à remercier Gopal Prasad pour l'aide qu'il m'a apportée, il y a plus de quarante ans, dans la détermination des corps constructibles.

Bibliographie

[Altınel et al. 2008] T. Altınel, A. V. Borovik et G. Cherlin, *Simple groups of finite Morley rank*, Mathematical Surveys and Monographs **145**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Ax 1968] J. Ax, "The elementary theory of finite fields", *Ann. of Math.* (2) **88** (1968), 239–271. MR Zbl

[Baldwin 1973] J. T. Baldwin, "α_T is finite for $\aleph_1$-categorical T", *Trans. Amer. Math. Soc.* **181** (1973), 37–51. MR Zbl

[Baudisch 1996] A. Baudisch, "A new uncountably categorical group", *Trans. Amer. Math. Soc.* **348**:10 (1996), 3889–3940. MR Zbl

[Baudisch et al. 2009] A. Baudisch, M. Hils, A. Martin-Pizarro et F. O. Wagner, "Die böse Farbe", *J. Inst. Math. Jussieu* **8**:3 (2009), 415–443. MR Zbl

[Belegradek 1973] O. V. Belegradek, "Almost categorical theories", *Sibirsk. Mat. Zh.* **14** (1973), 277–288. MR Zbl

[Berline 1986] C. Berline, "Superstable groups; a partial answer to conjectures of Cherlin and Zilber", *Ann. Pure Appl. Logic* **30**:1 (1986), 45–61. MR Zbl

[Berline et Lascar 1986] C. Berline et D. Lascar, "Superstable groups", *Ann. Pure Appl. Logic* **30**:1 (1986), 1–43. MR Zbl

[Blum 1969] L. C. Blum, *Generalized algebraic theories: a model theoretic approach*, thèse de doctorat, Massachusetts Institute of Technology, 1969, voir https://www.proquest.com/docview/302381589. MR

[Borovik 1984a] A. V. Borovik, "Инволюции в группах с размерностью [Involutions dans les groupes avec dimension]", prépublication 84-512, Centre de calcul de la division sibérienne de l'académie des sciences de l'URSS, Novosibirsk, 1984. MR

[Borovik 1984b] A. V. Borovik, "Теория конечных групп и несчётно категоричные группы [Théorie des groupes finis et groupes incomptablement catégoriques]", prépublication 84-511, Centre de calcul de la division sibérienne de l'académie des sciences de l'URSS, Novosibirsk, 1984. MR

[Borovik et Nesin 1994] A. Borovik et A. Nesin, *Groups of finite Morley rank*, Oxford Logic Guides **26**, The Clarendon Press, New York, 1994. MR Zbl

[Borovik et Poizat 1990] A. V. Borovik et B. P. Poizat, "Tores et p-groupes", *J. Symbolic Logic* **55**:2 (1990), 478–491. MR Zbl

[Burdges et Cherlin 2002] J. Burdges et G. Cherlin, "Borovik–Poizat rank and stability", *J. Symbolic Logic* **67**:4 (2002), 1570–1578. MR Zbl

[Cherlin 1978] G. Cherlin, "Superstable division rings", pp. 99–111 dans *Logic Colloquium '77* (Wrocław, 1977), édité par A. Macintyre et al., Stud. Logic Found. Math. **96**, North-Holland, 1978. MR Zbl

[Cherlin 1979] G. Cherlin, "Groups of small Morley rank", *Ann. Math. Logic* **17**:1-2 (1979), 1–28. MR Zbl

[Cherlin et Shelah 1980] G. Cherlin et S. Shelah, "Superstable fields and groups", *Ann. Math. Logic* **18**:3 (1980), 227–270. MR Zbl

[Clausen et Tent 2023] T. Clausen et K. Tent, "Mock hyperbolic reflection spaces and Frobenius groups of finite Morley rank", *Model Theory* **2**:2 (2023), 137–175. MR

[Deloro 2007] A. Deloro, *Groupes simples connexes minimaux de type impair*, thèse de doctorat, Université Denis-Diderot (Paris-VII), 2007, voir https://theses.fr/2007PA077036.

[van den Dries 1982] L. P. D. van den Dries, "Definable groups in characteristic 0 are algebraic groups", *Abstracts Amer. Math. Soc.* 3 (1982), 142.

[Frécon 2018] O. Frécon, "Simple groups of Morley rank 3 are algebraic", *J. Amer. Math. Soc.* **31**:3 (2018), 643–659. MR Zbl

[Glauberman 1966] G. Glauberman, "Central elements in core-free groups", *J. Algebra* **4** (1966), 403–420. MR Zbl

[Grishkov et Nagy 2011] A. Grishkov et G. P. Nagy, "Algebraic Bol loops", *Forum Math.* **23**:3 (2011), 655–668. MR Zbl

[Hrushovski 1986] E. Hrushovski, *Contributions to stable model theory*, thèse de doctorat, University of California, Berkeley, 1986, voir https://www.proquest.com/docview/303450288.

[Hrushovski 1992] E. Hrushovski, "Strongly minimal expansions of algebraically closed fields", *Israel J. Math.* **79**:2-3 (1992), 129–151. MR Zbl

[Hrushovski et Zilber 1996] E. Hrushovski et B. Zilber, "Zariski geometries", *J. Amer. Math. Soc.* **9**:1 (1996), 1–56. MR Zbl

[Kramer et al. 1999] L. Kramer, G. Röhrle et K. Tent, "Defining k in $G(k)$", *J. Algebra* **216**:1 (1999), 77–85. MR Zbl

[Lachlan 1980] A. H. Lachlan, "Singular properties of Morley rank", *Fund. Math.* **108**:3 (1980), 145–157. MR Zbl

[Lascar 1976] D. Lascar, "Ranks and definability in superstable theories", *Israel J. Math.* **23**:1 (1976), 53–87. MR Zbl

[Lascar 1985] D. Lascar, "Les groupes ω-stables de rang fini", *Trans. Amer. Math. Soc.* **292**:2 (1985), 451–462. MR Zbl

[Macintyre 1971] A. Macintyre, "On ω_1-categorical theories of fields", *Fund. Math.* **71**:1 (1971), 1–25. MR Zbl

[Macpherson et Pillay 1995] D. Macpherson et A. Pillay, "Primitive permutation groups of finite Morley rank", *Proc. London Math. Soc.* (3) **70**:3 (1995), 481–504. MR Zbl

[Maltsev 1940] A. Maltsev, "On isomorphic matrix representations of infinite groups", *Rec. Math. Moscou, N.S.* **8**:50 (1940), 405–422. En russe. MR Zbl

[Mustafin et Poizat 2006] Y. Mustafin et B. Poizat, "Sous-groupes superstables de $\mathrm{SL}_2(K)$ et de $\mathrm{PSL}_2(K)$", *J. Algebra* **297**:1 (2006), 155–167. MR Zbl

[Ould Houcine 2007] A. Ould Houcine, "On superstable groups with residual properties", *MLQ Math. Log. Q.* **53**:1 (2007), 19–26. MR Zbl

[Pillay 2013] A. Pillay, "Review of *Zariski geometries: Geometry from the logician's point of view* by Boris Zilber", *Bull. Amer. Math. Soc.* (*N.S.*) **50**:1 (2013), 175–180. MR

[Poizat 1978] B. Poizat, "Une preuve par la théorie de la déviation d'un théorème de J. Baldwin", *C. R. Acad. Sci. Paris Sér. A* **287**:8 (1978), 589–591. MR Zbl

[Poizat 1987] B. Poizat, *Groupes stables : une tentative de conciliation entre la géométrie algébrique et la logique mathématique*, Nur al-Mantiq wal-Ma'rifah **2**, Bruno Poizat, Lyon, 1987. MR Zbl

[Poizat 2000] B. Poizat, "Autour du théorème de Morley", pp. 879–896 dans *Development of mathematics* 1950–2000, édité par J.-P. Pier, Birkhäuser, 2000. MR Zbl

[Poizat 2001a] B. Poizat, "L'égalité au cube", *J. Symbolic Logic* **66**:4 (2001), 1647–1676. MR Zbl

[Poizat 2001b] B. Poizat, "Quelques modestes remarques à propos d'une conséquence inattendue d'un résultat surprenant de Monsieur Frank Olaf Wagner", *J. Symbolic Logic* **66**:4 (2001), 1637–1646. MR Zbl

[Poizat 2018] B. Poizat, "Milieu et symétrie, une étude de la convexité dans les groupes sans involutions", *J. Algebra* **497** (2018), 143–163. MR Zbl

[Poizat 2021] B. Poizat, "Symétries et transvexions, principalement dans les groupes de rang de Morley fini sans involutions", *J. Symb. Log.* **86**:3 (2021), 965–990. MR Zbl

[Poizat 2024] B. Poizat, "Quelques modestes compléments aux travaux de messieurs Mark Debonis, Franz Delahan, David Epstein et Ali Nesin sur les groupes de Frobenius de rang de Morley fini", *Model Theory* **3**:1 (2024), 71–100.

[Poizat et Wagner 1993] B. Poizat et F. Wagner, "Sous-groupes périodiques d'un groupe stable", *J. Symbolic Logic* **58**:2 (1993), 385–400. MR Zbl

[Reineke 1975] J. Reineke, "Minimale Gruppen", *Z. Math. Logik Grundlagen Math.* **21**:4 (1975), 357–359. MR Zbl

[Sacks 1972] G. E. Sacks, *Saturated model theory*, W. A. Benjamin, Reading, MA, 1972. MR Zbl

[Thomas 1983] S. Thomas, "The classification of the simple periodic linear groups", *Arch. Math.* (*Basel*) **41**:2 (1983), 103–116. MR Zbl

[Wagner 2000a] F. O. Wagner, "Minimal fields", *J. Symbolic Logic* **65**:4 (2000), 1833–1835. MR Zbl

[Wagner 2000b] F. O. Wagner, *Simple theories*, Mathematics and its Applications **503**, Kluwer, Dordrecht, 2000. MR Zbl

[Wagner 2001] F. Wagner, "Fields of finite Morley rank", *J. Symbolic Logic* **66**:2 (2001), 703–706. MR Zbl

[Wagner 2003] F. O. Wagner, "Bad fields in positive characteristic", *Bull. London Math. Soc.* **35**:4 (2003), 499–502. MR Zbl

[Weil 1955] A. Weil, "On algebraic groups of transformations", *Amer. J. Math.* **77** (1955), 355–391. MR Zbl

[Zamour 2022] S. Zamour, *Étude des quasi-groupes de Frobenius et des K-boucles de rang de Morley fini*, thèse de doctorat, Université Claude Bernard (Lyon 1), 2022, voir https://theses.hal.science/tel-03765160.

[Zilber 1974] B. I. Zilber, "The transcendental rank of the formulas of an $\aleph_1$-categorical theory", *Mat. Zametki* **15** (1974), 321–329. En russe ; traduit en anglais à *Math. Notes Acad. Sci. USSR* **15**(2) (1974), 182–186. MR Zbl

[Zilber 1977] B. I. Zilber, "Groups and rings whose theory is categorical", *Fund. Math.* **95**:3 (1977), 173–188. En russe. MR Zbl

Received 13 Dec 2022. Revised 13 Mar 2023.

BRUNO POIZAT:

poizat@math.univ-lyon1.fr
Institut Camille Jordan, Université Claude Bernard (Lyon 1), Villeurbanne, France

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.505

Around the algebraicity problem in odd type

Gregory Cherlin

We discuss the algebraicity conjecture in odd type, with particular attention to some unfinished business involving work of Jeffrey Burdges.

The notion of strongly minimal set was known to Vaught. He, and probably others, knew the Steinitz theorem could be generalized.

Bill Marsh, 1966 [43]

Introduction	505
1. Classification in finite 2-rank	509
2. Uniqueness cases	516
3. Strong embedding	518
4. High Prüfer 2-rank	521
5. Prüfer 2-rank 2, 2-Rank at least 3	524
6. Lately: linearization theorems	525
Appendix A. More on strong embedding	526
Appendix B. Glossary	529
References	536

Introduction

According to my recollection, I first encountered the internal geometry of strongly minimal sets in Marsh's 1966 thesis. On looking back at that thesis, I find that Marsh indulges in very little speculation about that geometry, but at the time it seemed suggestive. Fortunately, the matter was not left there, and once the dust had settled and the mists had cleared, we found ourselves with a robust geometrical stability theory which supports applications. As a result, the distinction between pure and applied model theory has become less fraught than it once was. As

Research supported over the years by a series of NSF Grants, including DMS-1362974.
MSC2020: primary 03C60, 20E32; secondary 03-02, 20A15.
Keywords: group, Morley rank, simple group, odd type, Zilber, strong embedding, algebraic group, classification, unipotence, Burdges, L-star, algebra, model theory, algebraicity conjecture, Cherlin–Zilber, stability, model theoretic algebra.

Zilber anticipated, this development involves both the internal geometry of strongly minimal sets and a study of the groups interpretable therein.

Nonetheless, the algebraicity conjecture (or problem) remains unsettled: *are the simple groups of finite Morley rank algebraic*? At this point one leans toward the expectation that counterexamples exist — possibly, as Zilber has suggested, coming naturally from the general direction of analysis. Leaving all of that aside, I will address some unfinished business which is connected with a portion of the *Borovik program*. This has become a highly developed subject with a great deal of technical material inspired in part by finite group theory, in part by algebraic group theory, and occasionally by developments in pure model theory. The various sections of the glossary in Appendix B should be helpful as we get into the details, and may merit an early glance as well.

In another direction, there has been considerable progress in the direction of "linearization", which we touch on at the end — this is dealt with comprehensively in the contributions of Borovik and of Deloro to this volume [13; 29].

The Borovik program aims to do what can be done on the positive side of the problem with existing techniques, notably those modeled on methods of finite group theory, and to identify specific problematic configurations which resist such an analysis. This program has undergone three waves of development, as the power of existing techniques has been refined and their scope enlarged.

In the first instance, an extraneous "tameness" hypothesis[1] was liberally employed, in the manner of "stone soup". This amounts to listing *bad fields* as one of the known problematic configurations. In the second wave, the stone was removed from the soup and the focus turned to the group theoretic configurations associated with a hypothetical *minimal* counterexample. This is the K^* theory, described in detail below. In a third wave, we aim at somewhat more. This is the L^* theory.

The most striking achievement of the L^* theory is the proof of the algebraicity conjecture for simple groups of finite Morley rank having *infinite* 2-rank (that is, when there is an infinite elementary abelian 2-group present) [3]. I will be discussing some classification results in *finite* 2-rank, and, notably, some unpublished work of Burdges. Namely, Burdges was actively pursuing some ideas about L^* theory in finite 2-rank of a more technical character when he became distracted by other matters. Eventually I thought I should try to do something about that, so in January 2016, at his wedding breakfast, I ransacked Jeff's computer and made off with the relevant files. At this point, it seems high time to document the state of affairs of this material.

So here we are. The subject could certainly use a more comprehensive and systematic account of what has been learned on the side of finite 2-rank (which

[1] I.e., no fields were injured in the production of the group; cf. [11] and see also [46; 6; 51].

would include the case in which there are no involutions), and the underlying theory of torsion elements in groups of finite Morley rank; this theory developed relatively late. But here we confine ourselves to situations where Sylow 2-subgroups are not too small.[2]

Some time ago, in [15], reference was made to the possibility of "eventually …emancipating the odd type analysis from the K^*-hypothesis, a line of development which remains to be explored."[3] This was followed by the question ("Problem 1"):

> *Can one show that a simple group of finite Morley rank and degenerate type has no nontrivial involutory automorphism*?

We do not address this question, but we do take it seriously. We look in fact at how the theory would proceed in the presence of a positive solution to that problem (or a fair approximation to one).

We thank Adrien Deloro for very pertinent remarks concerning the contents of [28; 34].

A few remarks in the margin … [4]

Without going much further into the history of the subject — which I think is very interesting, but not my own concern here — I'll note that I don't think the algebraicity conjecture was particularly central to Zilber's own concerns (and not precisely my own either, though more so). From his side the trichotomy conjecture[5] seems central and the algebraicity conjecture could be taken as one expression of it with the particular virtue of being accessible to mathematicians generally. On the other hand, for a time at least — a critical time, perhaps — we were both under the influence of Macintyre's striking paper on ω_1-categorical fields [42] (I was in fact obsessed by that paper, myself, for several years), and it certainly points the way.[6]

In my own case, while working with Macintyre and several others with similar interests, I came across what to me seemed an intriguing notion of "connectedness" in Kegel and Wehrfritz's informative [39, p. 97] while thinking about "totally categorical" groups, namely:

> If X is any subgroup of G, put $X^0 = \bigcap \mathbf{C}_X Y$, where the intersection is taken over all subsets Y of G for which the index $|X : \mathbf{C}_X Y|$

[2] Cf. [44, 4th heading].

[3] Repeated in the notes to §II.6 of [3] in the following terms: "… *everything* we do depends on [L^*-theory]. A major open problem is to develop a parallel theory in odd type groups, at a comparable level of generality."

[4] As the referee remarks, this section has rather a large number of footnotes. We apologize in advance.

[5] Or perhaps more properly, dichotomy; cf. [38].

[6] Wikipedia as of November 2022: "… very influential in the development of geometric stability theory.[citation needed]"

> is finite. Clearly for every subgroup X of G the index $|X : X^0|$ is finite since G is an $\mathfrak{M}_c$-group. Call the subgroup X *connected* (*in* G) if $X = X^0$. Obviously A, $\mathbf{Z}G$ and G are all connected.

Much as in the case of Marsh's thesis, I suppose, this prompted some reveries that were perhaps not in the text.

And at some point fairly early on I became aware, one way or another, of Zilber's remarkable "ladder theorem", to the effect that all uncountably categorical structures are built from a strongly minimal set, the operation of algebraic closure, and some group actions by definable groups. This result would suggest that the algebraicity conjecture might be a major ingredient in the classification of the possible structures. The relatively recent work on permutation groups of finite Morley rank (noted by Borovik elsewhere in this volume) aims to address this to some extent. Around 1980, I became particularly interested in some possibilities for using the projected classification of the finite simple groups in model theory, which is really a rather different subject from the algebraicity conjecture, in terms of its aims and content — though compatible with the Borovik program. As the first algebra course I took was given by Walter Feit, and the first Janko group came along while I was an undergraduate — and as I eventually found myself employed at what was at the time the world headquarters of that classification program (or its management) — this was a natural line to fall in with.

Fortunately, the reader looking for a more balanced and informative account of developments around the algebraicity conjecture may consult the historical survey by Poizat in this volume [47] for a coherent account of the subject that shows quite precisely how (though perhaps not why) the subject emerged into print, as far as both Zilber and I were concerned. This should be supplemented by Hodges' account in [37] of how some of us in the "West[7]" became aware of some important aspects of Zilber's thinking. That account enters into some detail concerning what was taking place either prior to publication or independently of publication.[8] From my own perspective Zilber's "VINITI" report [53] that Hodges mentions was particularly central, and I regret that I have not seen it in the last 40 years — at some point, as I have recently realized, I lost track of my own copy. It would be good, I think, to locate a copy of that report and put it in the public record. The VINITI report had a wealth of material,[9] some of which I lectured on in the model theory year in

[7] Including Vancouver.

[8] In addition the concise historical remarks at the end of the introduction to [3] may become more illuminating when combined with these two accounts.

[9] According to my current recollection, for what it is worth, 83 pages (though this is perhaps a subjective reaction to the fact that it was in Russian).

Jerusalem in 1980–1981.[10] To acquire that document, one had to write directly to a certain address in Moscow, and in due course one received one's personal copy. But I have now drifted entirely away from my present subject, to which I return.[11]

1. Classification in finite 2-rank: the K^* and L^* theories in a top-down approach

This section lays out our current subject matter with more precision, specifying the results and problems we will be focusing on here. That is, we describe two flavors of the Borovik program (K^* and L^*) and the "top-down" approach which has been taken over — greatly simplified — from the practice of finite group theorists. The reader may find a preliminary quick review of the material of Section B2 helpful at this point, before entering into the substance of the discussion here.

1A. ***K^* and L^*.*** Following the finite group theorists, we call a group of finite Morley rank a *K-group* if all of its definable connected simple sections are algebraic groups.[12] And we call a group a K^*-group if the same applies to its *proper* definable connected simple sections.

A K^*-group is either a K-group or a minimal counterexample to the algebraicity conjecture. That conjecture can be phrased in these terms as follows: all K^*-groups are K-groups. One particular version of the Borovik program aims at bounding the 2-rank of any exceptions. This can be done, which is satisfying in its own way, but this does not limit the complexity of an arbitrary counterexample to the conjecture. It means only that any counterexample to the algebraicity conjecture involves a counterexample of low complexity as a definable section.

It would be very valuable to have *absolute bounds* on the complexity of counterexamples. For that matter, even in the finite case one might well ask for a more qualitative proof that the number of sporadic finite simple groups is finite, or at least that their 2-ranks are bounded, without passing through an explicit classification of the exceptions.

The Borovik program (as such) focused on K^*-groups, sometimes with additional constraints. Altınel suggested a broader notion suitable for analyzing simple groups of finite Morley rank which have infinite 2-rank, essentially by relativizing the definition of K^* to this class.

[10] For some reason, what was actually on my own mind at the time was the problem of finiteness of Morley rank in the $\aleph_0$-categorical case, and the ideas of [41]. I had previously spent some effort in spring 1978 on the construction of an $\aleph_0$-categorical pseudoplane, without much success, a project I then abandoned.

[11] And I have managed to say not a word about the Soviet Union.

[12] This means, concretely, that these sections are Chevalley groups over algebraically closed fields — possibly with additional structure — and one will think mainly in those terms if one enters into the details. See Section B5.

More precisely, one first divides the groups of finite Morley rank of infinite 2-rank further into the following two classes:

(1) *even type* groups, where there is a bound on the exponents of 2-elements;

(2) *mixed type* groups, where there is no such bound, and in fact there is also a nontrivial *divisible* abelian 2-subgroup.

All algebraic groups of infinite 2-rank are of even type. Groups of mixed type and finite Morley rank do occur naturally, as products of algebraic groups (over different fields), but they do not occur as algebraic groups. In fact the finite Morley rank category is closed under finite direct products.

Definition 1.1 [3, Definition II.6.1]. A group of finite Morley rank and even type is an L-*group* if all of its definable simple sections of even type are algebraic.

L^*-*groups* (of even type) are defined correspondingly.

Here "L" stands for the first letter after "K", nothing more. The main theorems are the following, which we phrase as two independent results. We give references to [3], where one can also find historical notes and further bibliographic references; these results build on a large body of work with many contributors (among whom I feel one should take particular note of Éric Jaligot).

Proposition 1.2 [3, Proposition VIII.6.2]. *A simple L^*-group of finite Morley rank of even type is a Chevalley group.*

Proposition 1.3 [3, Mixed Type Theorem, Chapter V]. *If all simple groups of finite Morley rank and even type are algebraic, then there are no simple groups of finite Morley rank of mixed type.*

Putting the two propositions together, one has the following.

Theorem 1.4 [3, Main Theorem, Chapter X]. *A simple group of finite Morley rank and infinite 2-rank is algebraic.*

The striking and unexpected point is that Proposition 1.2 can be proved even though it makes no assumption on the degenerate type sections of the group in question. One proceeds largely by ignoring such sections — and, in particular, one does not worry at all about whether or not one might even enrich the structure of an algebraic group so as to make such a section appear. What makes this kind of analysis feasible is a sort of orthogonality principle.

Lemma 1.5 (Altınel's lemma, cf. [3, Proposition I.10.13]). *If a connected elementary abelian 2-group acts definably as a group of automorphisms of a connected group of finite Morley rank and finite 2-rank, then it acts trivially.*

This is the main reason to expect the K^* theory to go over in some form to an L^* theory. Lemma 1.5 can also be expressed in a structural form, as follows. The following is a slightly simplified formulation of [3, Proposition II.6.2].

Lemma 1.6. *Let G be an L-group of finite Morley rank and even type. Let $U_2(G)$ be the subgroup of G generated by its definable connected 2-subgroups. Then $U_2(G)$ is a K-group.*

Proofs in the L^* context are often more laborious than those in the K^* context, involving more exotic configurations, but the general approach taken is much the same under either hypothesis. The K^* project was still very much underway in the even type context when the L^* project came along, but after retracing its initial steps at this level of generality — with a more elaborate treatment of some uniqueness cases (strong embedding, weak embedding) — we switched over to that greater level of generality and finished the version of the classification project appropriate to infinite 2-rank in that setting [3]. As one might expect, another feature of that proof that it shares with the finite case is that it was the product of a community (including some members of the community that dealt with the finite case, who pointed out relevant strategic options not always leaping to the eye in the literature).

In view of Theorem 1.4, we may turn our attention to the case of finite 2-rank.

1B. ***Finite 2-rank: Prüfer and normal 2-ranks.*** In addition to the ordinary 2-rank, we have the important notion of *Prüfer* 2*-rank*.

Definition 1.7. Let G be a group of finite Morley rank. The 2-rank of a maximal 2-torus[13] in G is called its *Prüfer* 2*-rank*.

By a conjugacy theorem, this notion is well-defined. The Prüfer 2-rank corresponds to the Lie rank in the relevant groups (where the base field is algebraically closed and the characteristic is not 2). The Prüfer 2-rank is essential for our purposes. Groups of finite Morley rank and finite 2-rank are divided into the following types, again following the lead and terminology of the finite group theorists (though not, in this instance, their definitions).

- Prüfer 2-rank 0: *degenerate type*.
- Prüfer 2-rank 1: *thin type*.
- Prüfer 2-rank 2: *quasithin type*.
- Prüfer 2-rank at least 3: *generic type*.

The experience of finite group theorists suggests that the high end of the problem should be the most amenable to systematic treatment, and that the complexity and the general weirdness of the analysis will increase as one moves downward from the top.

Groups of degenerate type have 2-rank 0 (i.e., no involutions at all). In fact, in a group of finite Morley rank and finite 2-rank, any involution belongs to some 2-torus [20]. The nondegenerate type groups of finite Morley rank and finite 2-rank

[13] Divisible abelian 2-subgroup; a product of “Prüfer 2-groups,” which are 2-tori of 2-rank 1.

are said to be of *odd type*, since the algebraic ones have base fields of characteristic not two (so even 0 is “odd”, oddly enough). And all nonzero Prüfer 2-ranks (i.e., Lie ranks) certainly occur.

There are other ways of organizing this family of groups that do not match up neatly with the above taxonomy — at least, not a priori. One encounters the theoretical possibility of *minimal connected simple* groups regardless of Prüfer rank, and notably the so-called *uniqueness cases* which will be of considerable importance. But we will come back to that later.

Our focus here will be on the quasithin and generic type cases, that is, the Prüfer 2-rank is taken to be at least 2.

There is yet another notion of 2-rank which plays a role similar to that of the Prüfer 2-rank in the finite case, namely the normal 2-rank. This is defined as follows.

Definition 1.8. The maximal rank of a normal elementary abelian 2-subgroup of a Sylow 2-subgroup of G is called its *normal* 2*-rank*.

Again, by a conjugacy theorem, this is well-defined. In the cases of interest to us here, this parameter actually agrees with the Prüfer 2-rank. This took some time to be noticed, and is nontrivial.

Fact 1.9 [28, Lemma 1]. *Let G be a connected group of finite Morley rank and finite* 2*-rank. Then the Prüfer* 2*-rank and normal* 2*-rank of G coincide.*

1C. *L* revisited.* What happens if we relativize the notion of K^*-group to the class of *odd type* groups, rather than even type?

Definition 1.10. A group of odd type is called an *L-group* if all of its definable connected simple sections of odd type are Chevalley groups. We define *L^*-groups* of odd type correspondingly.

There is an immediate obstacle to the development of the theory: the lack of a known analog for Altınel’s lemma in this context. Such a lemma would control actions of 2-tori on degenerate sections of groups of odd type. The natural action of the multiplicative group of a field of characteristic zero on its additive group is just such an action, so a blanket prohibition on them is out of the question in this context.

Burdges suggested, nonetheless, pursuing the version of the theory in odd type that imposes both the L^* hypothesis and a suitable analog of Altınel’s lemma. We will adopt the following terminology.

Definition 1.11. A group of odd type is said to be NTA_2 if whenever a definable section H_1 acts definably on a definable simple section H_2 of degenerate type, then any 2-torus in H_1 acts trivially on H_2. In other words, if the action is required to be faithful then H_1 must be of degenerate type.

One may read “NTA_2” as “no definable 2-toral actions”.

When we refer to the L^* theory in odd type we will generally be taking NTA_2 as an assumption as well. Conditions of this kind, and more generally the study of involutory automorphisms of simple groups of degenerate type, are certainly of interest, and have been considered in the literature.[14]

My sense — after working with the concept for a while — is that the combination of L^* and NTA_2 in odd type does correspond quite neatly to the point of departure for the L^* theory in even type, and that to the extent that one can work in that setting, the results are more informative than the results of the K^* theory.

That is, in odd type, the L^* hypothesis gives us a useful dividing line, separating issues belonging properly to the study of simple groups of degenerate type from those bearing directly on odd type. And while the weight of opinion no doubt favors the existence of degenerate type simple groups, the question of 2-toral actions on such groups appears to be more delicate.

As a technical remark on the definitions, Burdges noted that in addition to the critical property NTA_2, there could be significant issues with the Glauberman Z^* theorem.[15] But experience to date suggests configurations of this type can be eliminated by close analysis on an ad hoc basis. We will next review and compare the status (in the odd type setting) of the K^* theory on the one hand, and the L^* theory with NTA_2 on the other.

1D. ***K* theory in odd type: results.*** As far as published results are concerned, one has mainly the *generic case for K^*-groups.*

Theorem 1.12 [18]. *Let G be a simple K^*-group of finite Morley rank, finite 2-rank, and of generic type. Then G is algebraic.*

The proof follows the template developed in [7] or [8], with some technical improvements.

This proof forks at a very early stage, with one branch leading to the desired identification and the other branch leading off in an entirely different direction, arriving eventually at a contradiction. This point is of central importance, so we give more detail.

[14]Cf. [34], [30, §1.3], and (as previously mentioned) [15, Problem 1]. From a different direction, similar questions arise in connection with questions about actions of finite groups which arise in the theory of permutation groups of finite Morley rank. In particular a conjecture from [27] concerning actions of Alt_n or Sym_n leads fairly rapidly to consideration of such actions on simple groups of degenerate type, as discussed in [2]. This direction might also provide some tightly constrained “extremal configurations” deserving close attention.

[15]See, for example, https://en.wikipedia.org/w/index.php?title=Z∗_theorem&oldid=1095862232. The original proof uses character theory. A proof which may well be of more use in the setting of finite Morley rank is in [52]; I have not looked into that, but it seems well worth looking at. (One can run into difficulties with arguments of an elementary nature as well; for example, the easy group theoretic proof that a group of order $2m$ with m odd has a subgroup of index 2 does not go over very readily to the finite Morley rank context.)

Definition 1.13. Let G be a group of finite Morley rank and S a 2-subgroup. Then $\Gamma_{S,2}(G)$ denotes the smallest definable subgroup of G containing the normalizer of every elementary abelian subgroup of S of rank 2.

The 2-*generated core* of G is $\Gamma_{S,2}(G)$ with S a Sylow 2-subgroup, which is well-defined up to conjugacy.

The case division that concerns us is whether or not the 2-generated core of G is proper. If it is, we are in a somewhat exceptional situation of the type referred to generally as a "uniqueness" case or "black hole," where it is hard to get at the whole of G using local analysis. In these cases we aim to push the configuration steadily to more extreme forms, the most extreme such case (short of a contradiction) being *strong embedding*: here some proper definable subgroup with an involution contains the normalizers of all of its nontrivial 2-subgroups.

On the main line of the proof, Theorem 1.12 takes the form that one has either a proper 2-generated core or one arrives at the desired conclusion.

On the uniqueness branch of the analysis, one has the following.

Theorem 1.14. *Let G be a simple K^*-group of finite Morley rank and odd type with a proper* 2*-generated core* $M = \Gamma_{S,2}(G)$.

(1) *If G has Prüfer* 2*-rank at least* 2, *and normal* 2*-rank at least* 3, *then M is strongly embedded and G is a minimal connected simple group.*

(2) *If G is minimal connected simple*, *then G has Prüfer* 2*-rank* 1.

These two points are dealt with in [17] and [26], respectively, and jointly eliminate this branch of the analysis in the generic setting, though the second point is of continuing interest. Burdges' involvement here indicates, among other things, that this is (at least) the second iteration of the Borovik program, and that once again there are major precursors to these results, under less general conditions.

Turning to the case of Prüfer rank 2, we have three types of algebraic groups to identify, of 2-rank at most 4, as shown in Table 1.

Burdges' work on this problem is unpublished even in the K^* setting, and is thoroughly entangled with the development of the L^* theory, so we will discuss it in that context. His treatment of 2-rank at least 4 was complete (pending further review of the details) and we will say more about that below. The treatment of 2-rank 3 led to an interesting exotic configuration similar to G_2 in characteristic 3 and known to the finite group theorists as worthy of separate consideration. In the finite case, they eliminated this exotic configuration via character theory (at first

type	A_2 (PSL_3)	G_2	B_2 (PSp_4)
2-*rank*	2	3	4

Table 1. Algebraic groups of Lie rank 2.

using modular character theory, and later by ordinary character theory). There is also prior work in 2-rank 2 by Altseimer and there were hopes of returning to that, but we are still dealing with the more accessible cases of higher 2-rank.

1E. *L^* theory in odd type: results.* Recall that here we will always be assuming the condition NTA_2. We turn to mostly unpublished results. There are some useful partial results that did make it into print at this level of generality, and some hints of the general program can be seen.[16] Here I give my views of what is currently known.

One result that is in print concerns the case of strong embedding (Theorem 2.2 below). A simplified version of that is the following.

Fact 1.15. *Let G be a simple L^*-group of finite Morley rank and finite 2-rank, with a strongly embedded subgroup, and Prüfer 2-rank at least 2. Then the 2-rank of G is its Prüfer 2-rank.*

This result (and the full version given later) is less satisfactory than what we have in the K^* case, where the Prüfer rank is reduced to 1, and the situation merits further exploration. Thus the uniqueness branch of the L^* theory remains open in general, and our statements of classification results reflect that.

In particular, for the generic case, the classification result reads as follows.

Theorem 1.16 (Burdges, cf. [21]). *A simple NTA_2 L^*-group of finite Morley rank and Prüfer 2-rank at least 3 is either a Chevalley group or has a strongly embedded subgroup (so in the latter case, it has 2-rank equal to its Prüfer rank).*

Coming to the case of Prüfer rank 2, the results currently focus on the cases with 2-rank at least 3. Here the uniqueness case does not pose difficulties (if one uses the full force of Theorem 2.2). But other problems arise in the case of 2-rank 3.

Theorem 1.17 [22; 23]. *A simple NTA_2 L^*-group of finite Morley rank, Prüfer 2-rank 2, and 2-rank at least 4 is PSp_4.*

For the case of 2-rank 3 the result claimed is as follows.

Theorem 1.18 [22; 23; 24]. *A simple NTA_2 L^*-group G of finite Morley rank, Prüfer 2-rank 2, and 2-rank 3 satisfies one of the following conditions.*

(1) *$G \simeq G_2$ over an algebraically closed field.*

(2) *All involutions of G are conjugate; for i an involution of G, $C(i)$ has the form $\mathrm{SL}_3 * \mathrm{SL}_3$ (possibly over different fields); the characteristics of both base fields are 3, and their ranks are equal; and if $\mathbb{X}$ is the product of two root groups, one from each factor, then*

$$N_G^\circ(\mathbb{X}) \le C(i). \qquad (*)$$

[16] In [19, §0.5] one can find some thoughts about an L^* theory (with no mention of NTA_2). Some results on L^* groups in the same vein as the K^* results are given there.

Condition (∗) is the exotic condition. This seems to be the most elaborate configuration currently known which may possibly occur in a minimal counterexample to the algebraicity conjecture.

The preprints [21; 22; 23; 24] can be found (in some form) at my website and should become available at arXiv after some additional review and editing. As Burdges has become occupied with other matters it may take some time for that to advance.[17]

In principle one could try to push the analysis further to the level of Prüfer 2-rank 2 and 2-rank 2, but one loses the signalizer functor theory, and with it, the last general method of component analysis in centralizers of involutions. There is some prior unpublished work by Altseimer in the K^* setting. Beyond this point, one approaches the theories of minimal simple groups and groups of degenerate type. At that point, while one is not necessarily limited to the K^* setting, the L^* theory does not provide a useful point of view.

We spend the rest of the present article filling out our discussion of the results mentioned, with one further note at the end. Accordingly, our usual notation and hypotheses will be as follows.

Hypothesis 1.19. *G is a connected simple group of finite Morley rank and finite 2-rank. It is an L^* group and satisfies condition* NTA_2.

Of course, before dealing with the L^* case one also develops some theory for L-groups with NTA_2, mainly in the connected case.

2. Uniqueness cases

In this section and the next we return to a discussion of the first main branch in the various analyses under discussion, namely splitting off the so-called uniqueness cases. The reader who prefers to follow the other branch can pass on to Section 4 with no loss of continuity.

The starting point for the uniqueness case is Theorem 2.2 below, which like everything in this line follows on several iterations of similar results which have been proved under varying hypotheses. It makes use of the following key notion, and several other technical terms which will be discussed further.

[17] I mention in passing that it has occurred to me that the sections devoted to ‘background material” in those preprints might be helpful in thinking about the content and focus of a comprehensive introduction to the theory of groups of finite Morley rank in odd type, for which it seems a book would be considerably more useful than a survey, as the material needs to be revisited, unified, and put at the level of generality most suited for potential applications. In the preprints, what can be quoted is quoted, and what needs elaboration or variation is elaborated on, or varied, by ad hoc arguments, with respect to the precise needs of the main argument.

Definition 2.1. Let G be a group of finite Morley rank and V an elementary abelian 2-subgroup. Then $\Gamma_V(G)$ is the subgroup generated by $C_G^\circ(E)$ with E varying over subgroups of V of index 2.

We are interested in Γ_V for V elementary abelian of rank 2, and thus Γ_V is generated by connected components of centralizers of involutions. Some terminology used in the following will be explained afterward.

Theorem 2.2 (generation theorem: [19, Theorem 6.6][18]). *Let G be a simple L^*-group of finite Morley rank and odd type with 2-rank at least 3, and V an elementary abelian 2-group of rank 2 with $\Gamma_V < G$. Then the following hold.*

(1) *The normalizer of Γ_V in G is a strongly embedded definable subgroup.*

(2) *G is a D^*-group.*

(3) *The Sylow 2-subgroups of G are 2-tori.*

(4) *The Weyl group W is nontrivial, and if r is the smallest prime divisor of the order of the Weyl group, then G contains a nontrivial unipotent r-subgroup.*

In particular, if one starts with any strongly embedded subgroup M then for any elementary abelian subgroup V of 2-rank at least 2 one has $\Gamma_V \leq M$ and thus one arrives at the same conclusions, with a strongly embedded subgroup which is definable (under the stated assumptions on G).

Now let us look at the rest of the terminology used above. First of all, in the L^* context the analogs of solvable group and minimal connected simple group are *D-group* and *D^*-group*, respectively.

Definition 2.3. A connected group of finite Morley rank is a *D-group* if all of its connected simple definable sections are of degenerate type. *D^*-groups* are defined similarly in terms of proper sections.

This is not to say that the class of D-groups is a *satisfactory* generalization of the class of solvable groups, but it is the class we are forced to deal with in this context.

The second point concerns the Weyl group, defined classically (for compact Lie groups, and later for algebraic groups) as $N(T)/T$ for a maximal torus (in the appropriate sense). But here the Weyl group is defined somewhat differently.

Definition 2.4. Let G be a group of finite Morley rank. A *decent torus* in G is a definable divisible abelian subgroup which is the definable hull of its torsion subgroup.

[18] Note on [19]: The statement of Theorem 1.2 given there is over-enthusiastic in its level of generality, replacing 2 by an arbitrary prime p, but overlooking a step where in fact p should have been 2. The situation, and the appropriate level of generality for the various results, is clarified by the remarks in the corresponding MathSciNet review, and in more detail in [28, Section 4].

The *Weyl group* associated with a maximal decent torus T is the finite group

$$W_T = N(T)/C(T).$$

By a conjugacy theorem, this is well-defined up to conjugacy in G.

One ingredient in the proof of Theorem 2.2 is a generation theorem for groups of degenerate type.

Theorem 2.5 [15, Theorem 5]. *Let G be a group of finite Morley rank of degenerate type and V a 4-group acting definably on G. Then $G = \Gamma_V(G)$; that is, G is generated by the connected centralizers*

$$C_G^\circ(a) \quad (a \text{ an involution in } V).$$

It was at this point in [15] that the question of emancipating oneself from the K^* hypothesis was raised.

Returning to the last point of Theorem 2.2, we can be more explicit about the nontriviality of the Weyl group. One begins with a Sylow 2-subgroup S contained in the strongly embedded subgroup M. At this point this is known to be a 2-torus. In this situation the involutions of S are conjugate in M and by a Frattini argument W_S acts transitively on this set. So W_S is nontrivial. Now take a maximal decent torus T containing S. By a Frattini argument W_T induces W_S on S.

3. Strong embedding

We discuss the proof of Theorem 1.14(2) from the point of view of the L^* setting. One would like to prove that a D^*-group of finite Morley rank and odd type satisfying the condition NTA_2, and having a definable strongly embedded subgroup M, must have Prüfer 2-rank 1. In the K^* case one has a choice of proofs, following either [26] or [4, Theorem 6.1]. Here we will be following the latter, but work in part with D^*-groups with NTA_2 rather than with minimal connected simple groups. There is a great deal of additional material which may be relevant, found in (at least) the papers [1; 4; 5; 17; 19; 18; 20; 26; 34].

We begin as follows. I am now discussing material for which there appears to be no formal reference at this level of generality; but see [4] for a highly relevant discussion — although the setting for that discussion was K^* theory, it very likely was also intended at the time to serve as a partial template for an approach to the L^* setting.

Proposition 3.1. *Let G be a D^*-group of finite Morley rank and odd type satisfying the condition NTA_2, and suppose G has a definable strongly embedded subgroup M. Set $B = M^\circ$ and let T be a Sylow 2-subgroup of M, F the Fitting subgroup of B (the largest definable connected nilpotent subgroup). Then the following hold.*

(1) $T \cap F = 1$.

(2) *If the Prüfer 2-rank is at least 2 then $M > B$ and $|M/B|$ is odd.*

(3) *If the Prüfer* 2*-rank is at least* 2 *then* B *is a maximal definable connected subgroup of* G.

(4) B *is a generous subgroup of* G (*i.e.*, *a generic element of* G *belongs to a conjugate of* B).

There is a major case division in the proof according as the involutions of M lie inside the Fitting subgroup or outside it. Our first point above indicates that the first of these cases can be eliminated. The method in this case is to show that this would lead to two disjoint generic subsets of the connected group G. This line of argument is taken over from earlier analysis, but involves quite a bit of structural analysis. It is interesting to see how that plays out when we cannot immediately invoke the theory of solvable groups via the hypothesis of minimal connected simplicity.

The rest of the analysis above bears on the second case. The second point follows quickly from the first given that the involutions of T must be conjugate in M; we touched on this point earlier.

With Proposition 3.1 in hand, we return to the K^* context (and, specifically, to the last few lines of [4]). Then the group B is solvable, so it is a Borel subgroup of G, and one may apply the following theorem to conclude.

Theorem 3.2 [4, Theorem 3.12]. *Let* G *be a minimal connected simple group of finite Morley rank and* B *a nonnilpotent generous Borel subgroup. Then* B *is self-normalizing.*

This can usefully be broken down somewhat, as follows. We note that the Weyl group is variously defined as $N_G(T)/C_G(T)$ or $N_G(T)/C_G^\circ(T)$ for T a maximal decent torus of G, but for G connected the definitions agree, by [1, Theorem 1].

We will work with the following four results in the minimal connected simple case, and use them to give a direct treatment of the case of strong embedding.

Fact 3.3. *Let* G *be a minimal connected simple group of finite Morley rank with nontrivial Weyl group* $W = W_T$ *of odd order*, p *a prime divisor of the order*, *and let* a *be a* p*-element of* G *representing an element of order* p *in* W.

(1) *If* p *is a minimal prime divisor of the order of* W *then* $C(a)$ *contains a nontrivial* p*-unipotent subgroup* [20, Corollary 5.2].

(2) *If* G *is of degenerate type and* $C(a)$ *contains a nontrivial* p*-unipotent subgroup then* G *contains no divisible* p*-torsion* [25, Lemma 3.5].

(3) *Suppose that* G *is of degenerate type, that* B_T *is a Borel subgroup containing* $C(T)$, *that* a *normalizes* B_T, *and that* T *contains no* p*-torsion. Then* $B_T = C(T)$ *is nilpotent and* $C_{B_T}(a) = 1$ [25, Proposition 3.10].

(4) *If* B *is a Borel subgroup containing a nontrivial unipotent* p*-subgroup then* p *does not divide* $[N_G(B) : B]$ [1, Lemma 4.3].

Setting aside the degeneracy hypotheses occurring in the literature, for which the assumption that the Weyl group is of odd order should be sufficient, and coming to the case of groups of finite Morley rank of odd type with a strongly embedded subgroup M, and Prüfer rank at least 2 (to ensure a nontrivial Weyl group), one can apply these conditions successively to $B_T = M^\circ$, with p the smallest prime divisor of the order of the Weyl group; (1) and (2) permit the application of (3) to conclude that $B_T = F(B_T)$, contradicting Proposition 3.1(1).

We did not use (4) here, but it comes in to the proof of (3). We also did not use Theorem 3.2 as stated, either. The proof of that theorem makes use of the following.

Fact 3.4 [25, Corollary 3.6]. *Let G be a minimal connected simple group of degenerate type with a nontrivial Weyl group W and let p the smallest prime divisor of the order of the Weyl group. Then there is no p-divisible torsion in G.*

This in turn makes use of Fact 3.3(1).

There is a general discrepancy between the way these various principles are stated and the way they are applied. In the literature one separates out the degenerate type analysis from the odd type analysis rather sharply and in the case of odd type one uses, in particular, the existing theory of groups with strongly embedded subgroups. If one wants to redo that theory by other methods then one may want to borrow the theory of groups with nontrivial Weyl groups of odd order from its usual context of degenerate type. One must then check (and also, mention) that the results do not in fact require degeneracy.

This point only arises if one is attempting to redo (or generalize) the theory of strongly embedded subgroups without making use of results that rely on that result. In the existing literature, several intertwined articles appeared at the same time, with this point more relevant at some points than at others, and in particular [4] discusses the way its results relate to those of [25] and why it is necessary to avoid quoting certain prior results in their most general form.

Also noteworthy is the following result from [34], as well as its method of proof and the comments Frécon makes on that proof and its relation to prior work. (The paper [34] takes a rather different point of view — or point of departure — from the one adopted here, and in particular makes use of somewhat different definitions. But Frécon begins by discussing the relations among the various points of view available quite thoroughly, and in particular shows that the conventions adopted do not conflict. We pass over these issues here and refer the interested reader to Frécon's account.)

Fact 3.5 [34, Lemma 2.8]. *Let G be a minimal connected simple group of finite Morley rank, C a Carter subgroup and p the smallest prime divisor of the order of the Weyl group W_G. If C has a nontrivial p-element, then G is of odd type and W_G has even order.*

In particular, under these hypotheses, and supposing we have a nontrivial Weyl group of odd order, it follows that there is no nontrivial p-torus in G. This result was obtained about the same time, but a little later, than Fact 3.3, but as Frécon notes, the proof is more direct. We are presently in a context where this sort of nuance can be of great technical value. Frécon's rather concise discussion following his proof of Lemma 2.8 of some of the dependencies among different parts of the theory as they were developed is valuable and pertinent.

One point of interest, for us, is the way the theory of solvable groups comes in. This frequently involves one of the following two points: p-unipotent subgroups of solvable groups lie in the Fitting subgroups, and in the minimal connected simple case, each p-unipotent subgroup lies in a unique Borel. One can easily conceive of reasonable hypotheses on degenerate type groups which might allow this type of principle to be extended to D^*-groups.

On the other hand, the theory also makes some use on occasion of some other delicate points from the solvable theory: conjugacy of Carter subgroups, and the delicate Bender analysis, which amounts to a close study of the maximal intersections of pairs of Borel subgroups. The latter topic once more invokes the properties of p-unipotent subgroups, but also brings in the characteristic zero unipotence theory. There are some results of striking generality on the theory of Carter subgroups of general groups of finite Morley rank, something one does not have in the finite case, not limited to the K^* case, but these do not fully cover the degenerate case. (See Section B7.)

One interesting question (with a great number of reasonable variants) is whether the mere assumption that B is solvable allows a similar treatment of the L^* case (which, recall, is just the D^* case at this point). In this form, using the existing techniques, this does not seem very likely, at least not without considerable additional work. We will continue this discussion, which is largely a discussion of open questions, in an appendix of a more exploratory character.

One could ask quite similar questions about the theory of Carter subgroups in groups of degenerate type, but we have not taken this up.

4. High Prüfer 2-rank

If we set aside the strongly embedded case, the identification of L^*-groups with NTA_2 in Prüfer 2-rank at least 3 may be completed. This is Theorem 1.16. In [18] an axiomatic framework for the proof was set out which is sufficient for the application of the argument of [7]. This framework is the following.

Hypothesis 4.1. *G is a connected simple group of finite Morley rank and odd type with Prüfer 2-rank at least 3.*

T_2 is a maximal 2-torus of G.

Σ is a family of subgroups of G of type (P)SL$_2$. We suppose that Σ has the following properties.

(1) $\Sigma^g = \Sigma$.

(2) $\langle \bigcup \Sigma \rangle = G$.

(3) For K in Σ we have

  (a) K is normalized by T_2.

  (b) $C_K(T_2)$ is a maximal algebraic torus of K; this torus is denoted $\mathbb{T}_K$.

  (c) $K = E(C_G(C_{T_2}(K)))$.

  (d) K is a Zariski closed subgroup of any definable algebraic quasisimple subgroup of G which contains K, and which is normalized by T_2.

(4) For K_1, K_2 in Σ distinct, and $L = \langle K_1, K_2 \rangle$, we have

  (a) $C_T(K_1) \cap C_T(K_2) \neq 1$.

  (b) Either K_1 and K_2 commute or L is an algebraic group of type A_2, $B_2 = C_2$, or G_2, and in that case K_1 and K_2 are root SL$_2$-subgroups of L normalized by $\mathbb{T}_L$.

  (c) The maximal tori $\mathbb{T}_{K_1}$, $\mathbb{T}_{K_2}$ associated with K_1 and K_2 commute.

  (d) $T_2 \cap L = (T_2 \cap K_1) * (T_2 \cap K_2)$ is a Sylow° 2-subgroup of L.

Such a family Σ is called a family of *root* SL$_2$-*subgroups* with respect to the 2-torus T_2.

The key to the construction of a suitable family Σ in the context of Theorem 1.16 is the identification of suitable algebraic subgroups of $C_G(i)$ for involutions $i \in T_2$. More precisely, one works within algebraic components of $EC_G(i)$.[19] Much of the work goes into the proof of the existence of these components.

Once one has sufficiently many such subgroups (as expressed by condition (2)), the argument becomes comparatively formal. The failure of the generation condition (2) leads to a strongly embedded subgroup. After that, the rest of the analysis involves ordinary root subgroups inside algebraic components L, taken with respect to the algebraic torus $\mathbb{T}_L$. One occasionally invokes the assumption on high Prüfer rank, and induction (4(a)), to bypass what would otherwise be challenging technical issues.

To reach the point of departure, namely the existence of algebraic components in centralizers of involutions, one needs *signalizer functor theory*. In this area, one quickly encounters difficulties associated with bad fields, or more particularly, a possible embedding of the additive group of one field in characteristic zero into the multiplicative group of another such field. This is overcome using Burdges' unipotence theory, and involves consideration of the reduced ranks of the fields involved, taking into account the generation of algebraic groups by unipotent subgroups.

[19] See Section B4.

As far as the classification of groups of high Prüfer rank is concerned, it is primarily the adaptation of the signalizer functor theory to the L^* setting that is still unpublished. For this, see the preprint [21]. We mention two noteworthy points.

First, one can put the hypothesis NTA_2 into a more directly applicable form. What follows is a slightly specialized version of what is currently Proposition 3.10 of [21] (in §3.3, *Structure of L-groups with* NTA_2).

Fact 4.2. *Let H be a connected L-group of finite Morley rank and odd type, satisfying NTA_2. Suppose that*

$$OF(H) \le Z(H).$$

Then

$$H = \mathrm{E}_{\mathrm{alg}}(H) * K \textit{ where } K \textit{ is connected and } K/Z^\circ(K) \textit{ has degenerate type.}$$

Here $\mathrm{E}_{\mathrm{alg}}$ *is the product of the* (*individually*) *algebraic components of $E(H)$.*

Hence the Sylow 2*-subgroup of K is central in H, and connected.*

This result is certainly useful as stated, but for technical reasons it seems necessary to give a sharper version, replacing $OF(H)$ by the largest connected normal subgroup of $F(H)$ whose torsion subgroup has bounded exponent.

Our second point, below, is particularly technical, but it captures one of the main points. The statement highlights the role of *cotorality* as well as some rank conditions relating to the Burdges unipotence theory. Furthermore, as formulated, the following can also be applied in some situations in Prüfer rank 2 and 2-rank at least 3.

Fact 4.3 [21, Lemma 4.4]. *Let G be an L^*-group of finite Morley rank of odd type, satisfying NTA_2. Let i, j, k be three commuting involutions in G and let ρ be either a prime or a symbol* $(0, r)$ *satisfying the conditions*

$$r > r_{f,i}, \qquad r \ge r_{0,i}.$$

Suppose the following.

(1) *i and j are cotoral in G.*

(2) $\theta_\rho(k) \cap C_G(j) \le \theta_\rho(j)$.

Then

$$\theta_\rho(k) \cap C_G(i) \le \theta_\rho(i).$$

Here two involutions are said to be *cotoral* if they lie in some 2-torus of the group G. It is known that in a connected group of odd type, each involution lies in some 2-torus. The same theory casts some light on the cotorality condition as well [20]. The issue of cotorality becomes more delicate, and therefore more important, in the context of Prüfer rank 2. But for some purposes the condition that the Prüfer 2-rank is at least 3 can be replaced by the much less restrictive condition there is an elementary abelian 2-group A of 2-rank 3 such that each pair of involutions in A is cotoral — or even somewhat less.

The use of the symbol ρ as either a prime p or a symbol $(0, r)$ refers to the Burdges unipotence theory, which extends the notion of p-unipotence to include a family of notions of $(0, r)$-unipotence associated with the "prime" 0. In particular the notation $\theta_\rho(k)$ is defined using the unipotence theory as the unipotent radical, in the sense of ρ, of the group $O^\sigma C(k)$ (the largest connected normal solvable definable subgroup of $C(k)$ without involutions). This is the sort of subgroup one always considers in connection with signalizer functor theory, for reasons touched on further in Section B4.

Noteworthy here is the restriction on ρ in terms of two associated parameters $r_{f,i}$ and $r_{0,i}$ involving the structure of $C_G^\circ(i)$. We say a bit more about these.

The parameter $r_{0,i}$ is familiar from Burdges unipotence theory as the maximal reduced rank associated with the odd solvable radical of $C_G(i)$. The parameter $r_{f,i}$ on the other hand is less often met with. Here the notation "f" stands for *field*, and the parameter $r_{f,i}$ measures the maximum reduced rank of the multiplicative group of a field which occurs as the base field of a component of the group

$$\mathrm{E}_{\mathrm{alg}}\big(C_G(i)/O^\sigma(C_G(i))\big),$$

where as above, O^σ means "largest connected normal solvable definable subgroup without involutions".

This analysis is finely tuned, and a good deal of it is foreshadowed by Burdges' thesis; cf. [18].

5. Prüfer 2-rank 2, 2-Rank at least 3

We come to the case of Prüfer 2-rank 2 and 2-rank at least 3. Here the strongly embedded configuration cannot arise — that would involve a connected Sylow 2-subgroup, and in that case the 2-rank and Prüfer 2-rank would coincide (Fact 1.15). So one may aim outright at identification.

As we explained at the end of Section 1D, in Prüfer 2-rank 2, even the K^* version of the results is unpublished.[20] We now resume this discussion once more, at the level of the L^* theory. This material is the subject of [22–24].

In 2-rank at least 4 one arrives at the expected identification: the group is PSp_4. In 2-rank 3 the target is G_2. In the course of the analysis in 2-rank 3 two cases arise, one leading to G_2 as expected. The other branch leads to a configuration quite familiar in the finite case, but eliminated there by character theory — initially, by modular character theory, and later by ordinary character theory. This exotic case is associated with a "base field" of characteristic three. It also has some uniqueness properties with respect to unipotent subgroups.

[20]And elaborate; and not, as yet, particularly closely vetted. I give my current view of it.

Apart from the exotic case, the treatments of the case of 2-rank at least 4 and the case of 2-rank 3 have a great deal in common. In very general terms, the line of analysis in both cases is similar to the analysis in high Prüfer rank, and in particular the signalizer functor theory is brought to bear in the same way. The first stage involves the analysis of algebraic components of centralizers of involutions, and the second stage focuses on the structure of the Weyl group. However, the treatment is far less "axiomatic" and involves much more detail at the level of particular configurations, with a number of undesirable configurations requiring close attention prior to their elimination.

The axiomatic approach in higher Prüfer rank was based on the idea that the Dynkin diagram encodes the structure of certain Lie rank 2 subgroups and in Prüfer rank 3 or more the induction hypothesis already controls the possibilities for these.

In the case of Prüfer rank 2, there is considerable uniformity in the treatment of components. When one brings in the Weyl group the two cases divide and each is handled separately and quite explicitly. Rather than applying general theory, the general thrust of the analysis is a direct examination of the action of the Weyl group on root subgroups and the verification of a qualitative form of the Chevalley commutator formula on a case-by-case basis.

The target ultimately is to invoke the theory of BN-pairs of finite Morley rank [40]. As we have mentioned, the first phase, having to do with the existence and precise determination of the algebraic components in centralizers of involutions, involves some close analysis of potentially pathological configurations. Here one has recourse to some specialized topics borrowed from finite group theory, notably the Thompson $A \times B$ theorem.

As far as the case of Prüfer rank 2 and 2-rank 2 is concerned, where one aims at identification of PSL_3, the terrain is largely unexplored, with the exception of early unpublished work by Altseimer. Here one should focus initially on the K^* context. In that case, one can at least rule out the strongly embedded configuration at the start.

6. Lately: linearization theorems

From the very beginning, understanding the representation theory of algebraic groups in the finite Morley rank category has been a major challenge, though partial results on the topic have already played a major role in such topics as the classification problem and the theory of permutation actions of finite Morley rank.

A recent milestone in this area is the following, a long-standing conjecture.[21]

[21] The genesis of this result, as a byproduct of measures taken in the recent (and as of this writing, ongoing) pandemic, is discussed in [12]. The author remarks "My triumph would be better deserved if it was someone else's conjecture; unfortunately, it was my own." This I take to be as much a sociological remark as a philosophical one.

Theorem 6.1 [13, Theorem 1.4]. *Let K be an algebraically closed field of characteristic $p > 0$ and G the group of points over K of a simple algebraic group defined over K. Assume that G acts definably and irreducibly on an elementary abelian p-group V of finite Morley rank. Then V has a structure of a finite dimensional K-vector space V_K, compatible with the action of G.*

The paper [13] uses enveloping algebra techniques in ways not previously seen in the subject. This result clearly has implications for the general study of definable actions of groups of finite Morley rank on abelian groups, and hence in the theory of permutation groups generally. See [10].

The theorem could well be of some relevance also in looking into some of the more recalcitrant configurations associated with the classification problem, such as the strongly embedded configuration in the context of L^* theory in odd type.

A comprehensive account will be found in Borovik's contribution to this volume [13]. In addition to the linearization of actions of algebraic groups, there is a more general form of linearization in the finite Morley rank context, which forces a field into existence. This "Schur–Zilber" approach is discussed by Deloro [29], in the context of a broad generalization of the result, and a novel approach.

Appendix A: More on strong embedding

As we have indicated, a satisfactory treatment of strong embedding relies on the hypothesis of minimal connected simplicity, that is, the assumption that all proper definable connected subgroups are solvable, and then on the theory of solvable groups of finite Morley rank, which is a rich subject that can be exploited in a variety of ways, notably via Carter subgroup theory.

More than one route has been taken to the treatment of this case in the K^* context. We indicated one such in the discussion around Fact 3.3.

Here we add a few comments about what is known in this direction more generally, and consider weakenings of minimal connected simplicity that suggest themselves as contexts for a broader treatment by the same methods.

A1. *A few general results.*

Fact A.1 [19, Theorem 6.6]. *Let G be a simple L^*-group of finite Morley rank of odd type, with $m_2(G) \geq 2$. Let V be an elementary abelian 2-group of rank 2 with $\Gamma_V < G$.*

If p is the least prime divisor of M/M°, then G contains a nontrivial unipotent q-subgroup for some prime $q \leq p$.

(One may rephrase the hypothesis on Γ_V as stating that G has a strongly embedded subgroup.)

Fact A.2 [15, Proposition 1.1, Theorem 4]. *Let G be a connected, nontrivial group of finite Morley rank and $g \in G$. Then the centralizer $C_G(g)$ is infinite. If g is a p-element for some prime p, then $C(g)$ contains an infinite abelian p-subgroup.*

Lemma A.3. *Let G be a D^*-group of finite Morley rank with a definable strongly embedded subgroup M, and $B = M^\circ$. Suppose also condition NTA_2.*

Let $w \in M \setminus B$. Then $C_B(w)$ is of unipotent type (i.e., contains no nontrivial divisible torsion).[22]

Proof. Suppose S is a nontrivial divisible abelian torsion subgroup in $C_B(w)$. Then $C_B(S)$ contains a maximal decent torus T of M and hence $C(S) \le M$. But $C(S)$ is connected [1, Theorem 1], and so $C(S) \le B$. Thus $w \in C(S) \le B$, a contradiction. □

Taking p to be a divisor of the order of the Weyl group in this setting, it follows that $U_p F(B) = 1$.

From the solvable theory, the following is key for our purposes. and stands apart from much of the rest of the theory.

Fact A.4 [25, Lemma 6.6 (Frécon)]. *Let G be a connected solvable group of finite Morley rank, and let $H < G$ be a definable connected subgroup of G such that $N_G^\circ(H) = H$. Then $N_G(H) = H$.*

It would be helpful to have something that can be put to similar use in the context of D-groups.

A2. *Generalizations of minimal connected simplicity.* The condition of minimal connected simplicity states that proper connected definable subgroups are solvable. One can generalize this condition either by restricting attention to some definable subgroups — notably $N^\circ(A)$ for suitable definable subgroups A — or by weakening the solvability condition, as in the D^*-condition.

Since there are substantial obstacles to the extension of the K^* theory to the D^* context it seems worthwhile to consider natural extensions of the K^* context which support the existing techniques. In that setting, one anticipates that the few points where the theory of solvable groups enters in a serious way might pose further problems, and if so it would be good to identify them.

We discuss some formal aspects of this. The following conditions are very natural.

Definition A.5. Let G be a group of finite Morley rank, p a prime, and H a definable subgroup.

[22] As remarked in [16], among all possible definitions of unipotence, this was "the broadest one we can imagine". Fortunately — given that — it is also a nontrivial condition that can be applied.

(1) The group G is *U_p-trivial* if for every degenerate type simple definable section L of G, any definable action of a unipotent p-group on L is trivial. In particular, L contains no nontrivial p-unipotent subgroup.

(2) H is *U_p-solvable* if the subgroup $U_p(H)$ generated by p-unipotent subgroups is solvable (equivalently, $U_p(H) \le F(H)$).

(3) A definable subgroup H of G is a *U_p-uniqueness* subgroup if for every nontrivial p-unipotent subgroup U of H, we have $N_H^\circ(U) \le H$; and we suppose also that H does in fact contain some nontrivial p-unipotent subgroup.

(4) G is *U_p-minimal* if every proper connected definable subgroup is U_p-solvable.

(5) G is an *$N_{U_p}^\circ$-group* if for every nontrivial abelian subgroup A the connected normalizer $N_G^\circ(A)$ is U_p-solvable.

Here condition (1) is a reasonable form of “tameness” to impose, allowing one to explore the configurations remaining when the more extreme configurations are eliminated. However one would not expect to work directly with that type of hypothesis, but rather with more abstract conditions of the type of (4) or (5). Here condition (3) is an expression of “Bender principle” used by Jaligot and developed further by Burdges, which one would expect to play a major role.

In particular solvable groups are U_p-solvable, minimal connected simple groups are U_p-minimal, and U_p-minimal connected simple groups are $N_{U_p}^\circ$-groups. And if a Borel subgroup B of a minimal connected simple group G contains a nontrivial p-unipotent subgroup, then it is a U_p-uniqueness subgroup.

We would also prefer to put more emphasis on the particular subgroup $B = M^\circ$, to the extent possible, and on the connected centralizers of Weyl group representatives, but not on the first pass.

Lemma A.6. *Let H be a connected D-group of finite Morley rank. If H is U_p-trivial then it is U_p-solvable.*

Proof. Let $U = U_p(H)$. By assumption $U \le C(EH)$.

We first treat the case in which

$$U_p F(H) = 1.$$

Then $U \le C(F(H))$ and hence $U \le C(F(H)E(H)) = Z(F(H))$, so $U = 1$ in this case.

For the general case, let $\overline{H} = H/U_p F(H)$. Then $U_p F(\overline{H}) = 1$ and $\overline{U} \le U_p(\overline{H}) = 1$, so $U \le U_p F(H)$ as claimed. □

Thus a connected simple D^*-group of finite Morley rank which is U_p-trivial is U_p-minimal.

Lemma A.7. *Let G be a connected simple D^*-group of finite Morley rank which is an $N^\circ_{U_p}$-group. Then any maximal connected definable U_p-solvable subgroup H with $U_p(H) > 1$ is a U_p-uniqueness subgroup.*

In other words, if we call a maximal connected definable U_p-solvable subgroup a U_p*-Borel* subgroup, then each U_p-local subgroup is contained in a unique U_p-Borel subgroup.

This is the usual argument.

Proof. Assuming the contrary we can find a nontrivial p-unipotent subgroup U of H and another maximal connected definable U_p-solvable subgroup H_1 containing it.

We may suppose further that the pair (H, H_1) is chosen to maximize U. If $U = U_p(H) = U_p(H_1)$ then $H = H_1$ for a contradiction, so we may suppose $U_p(H) > U$. Then $N^\circ_{U_p(H)}(U) > U$ and by maximality $N^\circ_G(U) \le H$. It follows that $U_p(H_1) \le H$, so $U = U_p(H_1)$ and $H_1 \le N^\circ_G(U) \le H$, for a contradiction. □

Appendix B: Glossary

Here we review some of the technical notions that occur at various points in the discussion, to simplify navigation. Many of these notions are not explicitly defined above, but in such cases we indicate the ideas behind them and the roles they play.

It may be helpful to have some of this collected together in one place. We also include some introductory remarks which are less directly pertinent to the technical discussion.

B1. ***Group theoretic terminology.*** We use terminology coming both from finite group theory and algebraic group theory. The one place where there is a notable terminological conflict between these two subjects is the use of the term "simple" in algebraic group theory in the sense of "quasisimple" in finite group theory. We use the term "simple" in its more literal sense.

Notions from finite group theory (or abstract group theory in general) can typically be taken over directly into our subject; notions from algebraic group theory may inspire similar notions with less geometrical definitions, which should be equivalent to the original definitions in the context of algebraic groups over algebraically closed fields *which carry no additional definable structure.* These notions are thoroughly covered by [14], and again, with some additions, in [3], where the bulk of Chapter I concerns various topics that fall under this heading.

We tend to work with *definable* subgroups, and definable sections (quotients of a definable subgroup by a definable normal subgroup). On the other hand, some very important subgroups that come into play tend not to be definable — notably, 2-Sylow subgroups — and accordingly when subgroups are meant to be definable, this is always specified. One has, in general, the *definable hull* of any subgroup —

or any subset — that is, the smallest definable group containing the given set. (One should avoid using the possibly more natural expression *definable closure* in this sense, as it has another meaning in model theory, of a very different character.)

It is a theorem of algebraic group theory that simple algebraic groups are linear groups, and the classification of the simple algebraic groups makes use of this point. We tend to identify these groups (which are functors) with the actual groups of rational points over an algebraically closed field. This is in some ways similar to talking about structures — or a particular structure — rather than theories. In a similar vein we tend to assume that our groups are saturated, though in the context of algebraic groups, there is occasionally a point to considering what happens over the algebraic closure of the prime subfield. None of this will be visible in the discussion in this paper; it lurks in the background.

Any algebraic group is a group of finite Morley rank, when realized concretely as a group over some fixed algebraically closed field, and the most striking applications of the theory to classical problems of mathematics, to date, actually come in the context of abelian varieties, and hence lie more or less at the opposite pole from the algebraicity conjecture. On the other hand, those applications pass in some cases through differential algebra, and in that context one has also a rich Galois theory and structural issues in the simple case, so in that respect at least, the subjects are not entirely foreign to one another. In this connection I would point to [35; 36].

Less concretely relevant, but I think of some importance, is the fact that the theory sits within the broader subject of *stable group theory*, which provides possibly the most satisfactory framework for thinking about the model theoretic issues that arise. For this the main point of entry remains [45], or its English translation.

In Table 2 (at the end) we list some more or less standard group theoretic operators whose definitions may vary a bit in the setting of groups of finite Morley rank, depending on whether or not issues of definability or connectedness arise. In most cases one proves definability under standard definitions. Notational conventions may vary, and we follow the preprints [21; 22] here, but most of this is found in [14].

Our convention here is that there is an ambient group G, and that H is one of the customary notations for a subgroup (more often than not, a definable subgroup). We mention that one occasionally takes connected components of nondefinable subgroups, using relatively definable subgroups (i.e., suitable intersections).

In the finite case the use of $O(G)$ is based on the Feit–Thompson theorem and it presumably corresponds more closely to $O^\sigma(G)$. The most immediate analog of "odd order" in the context of finite Morley rank is "without involutions". It is awkward to try to work with a very direct analog of the operator $O(G)$ as used in the finite case — consider for example an algebraic torus — so we pass directly to the connected analog. Most of the time it is used in a context like $OF(H)$ where it is already solvable, and one has the solvable version O^σ available when it is more appropriate.

B2. *Basic notions.* Simple groups of finite Morley rank are divided into *degenerate*, *odd*, *even*, and *mixed* types (pp. 511, 510).

The degenerate and odd types have finite 2-*rank* (zero or positive, respectively), while even and mixed type have infinite 2-rank (pp. 506, 511, 510). When the 2-rank is finite, a critical parameter is the *Prüfer* 2-rank, which corresponds in the algebraic setting to the Lie rank (p. 511).

The focus of the *Borovik program* (p. 506) has been on K^**-groups*, but it turns out that in even and mixed type, by extending the theory to the so-called L^**-groups* one can get a proof of the algebraicity conjecture for even and mixed type (p. 509).

A variant L^*-theory for odd type presents more difficulties. In this case the natural definition has to be supplemented by a condition denoted NTA_2 which is parallel to Altınel's Lemma 1.5 in the even or mixed type setting. The condition NTA_2 remains conjectural and appears to have roughly the level of difficulty of a full classification in the case of simple groups of Prüfer rank 1 (Definition 1.11).

In the L^*-theory one has also the more technical notions of D*-groups* and D^**-groups*, which play much the same role in that context as solvable groups and minimal connected simple groups do in the K^* context (Definition 2.3).

We subdivide odd type correspondingly into *thin*, *quasithin*, and *generic* type, corresponding to Prüfer 2-rank 1, 2, or higher (Definition 1.7).

We also must consider some notions of groups of *uniqueness* type, notably the case of *strong embedding*. One hopes that these groups will have Prüfer 2-rank 1. This is known in the K^* context (Theorems 1.14 and 2.2).

To get one's bearings in the technical side of the subject it is helpful to go to [18], which among other things provides a guide to a substantial body of prior work.

B3. *Torsion and Weyl groups.* A Π*-torus* is a divisible abelian torsion group and a p*-torus* is a divisible abelian p-group. A *decent torus* is the definable hull of a Π-torus. One has conjugacy theorems for the maximal p-tori, Π-tori, or decent tori. The *Weyl group* of a group G of finite Morley rank is the finite group $N_G(T)/C_G(T)$ where T is a maximal decent torus (or a maximal Π-torus). See Definition 2.4.

The study of torsion in groups of finite Morley rank leads into the close study of Weyl groups in exotic configurations and is of particular importance in groups which are small in the sense of uniqueness type or which are minimal connected simple (p. 519).

B4. *Classification techniques.* We have made rather cavalier use of the notation $EC(i)$, beginning with p. 522. This permeates the classification theory for finite groups as well as the L^* theory as discussed here, So we elaborate.

Here i is an involution, and $C(i)$ its centralizer. In an odd type group an involution plays the role of a semisimple element, and the conjecture we aim to prove predicts the structure of this group very precisely. The subgroup $EC(i)$ is the largest normal subgroup of $C(i)$ which is a direct product of quasisimple groups (that is, groups which are simple modulo their center, and perfect). For our purposes only the algebraic factors of $EC(i)$ are useful; the degenerate factors will be ignored. If one has enough algebraic factors then one hopes to reconstruct the entire group from them. Here the odd solvable radical $O^\sigma C_G^\circ(i)$ represents a potential obstacle to this.

The key ingredient in the analysis of $EC(i)$ (more specifically, for the control of $O^\sigma C(i)$) is the highly technical *signalizer functor* theory (p. 522). This leads eventually to the desired algebraic components of $EC(i)$.

This theory makes extensive use of the *Burdges unipotence theory*, which provides a characteristic zero analog of the *p-unipotence theory* (p. 523). The theory also requires a good understanding of *torality* and *cotorality* of involutions (p. 523), when one comes to the case of Prüfer 2-rank 2.

For our purposes, the most important point of the Burdges unipotence theory is that the additive group of a field is "more unipotent" than its multiplicative group (and also, that simple algebraic groups are generated by copies of additive groups of the base field). This feeds into the signalizer functor theory via a study of "sufficiently unipotent" base fields, in the case in which simple algebraic sections of a given group involve more than one base field. The precise measure of this is given by two parameters denoted $r_{f,i}$ and $r_{0,i}$ associated with an involution i, where the subscript "f" refers to base fields and the subscript "0" refers to the general unipotence theory in characteristic zero. We do not give further details here.

Another important feature of the unipotence theory is a notion of unipotent radical. Given that there are several notions of unipotence in play, there are several associated notions of unipotent radical, and not all are well-behaved. Subscripts as in θ_ρ tend to make (oblique) references to such notions.

B5. ***Simple algebraic groups.*** We rely in a certain sense on the classification of the simple algebraic groups as Chevalley groups over algebraically closed fields. The point here is that we have no hope of classifying the possible theories of these groups in an arbitrary language. There is presently a very rich supply of theories of algebraically closed fields of finite Morley rank, inspired by Hrushovski's refutation of the original formulation of Zilber trichotomy. We aim only to classify the groups obtained as abstract groups, and for this it is very natural to work toward some explicit presentation of the group; some such approach remains necessary to establish the existence of these groups, in fact, a point which remained an oddly open question for half a century between the construction of the smallest algebraic group of exceptional type by Dickson and Chevalley's explanation of how to use a suitably chosen basis for the associated Lie algebra as to allow for a sufficiently well-defined exponential map over an arbitrary base field. Standard references for this would be Steinberg's notes and Carter's book, both of which work very directly with explicit generators and relations. Chevalley had also asked for a more geometrical theory, and in addition to the general theory of algebraic groups, the Tits theory of buildings (and its later refinement to Moufang buildings) provides a satisfactory approach. On a more technical level, Tits' more ad hoc theory of BN-pairs presents an efficient way of reaching the Chevalley–Steinberg relations. Both of these approaches of Tits have been taken over to the context of groups of finite Morley rank and provide essential tools for the efficient recognition of simple groups as algebraic groups once a sufficient amount of group theoretic structural analysis has been carried out.

It will be noticed that in this context we are never actually in a position to use the classification of the simple algebraic groups, but it tells us what we are aiming for. The same is true in the context of finite group theory, where one has also to make one's way past 26 sporadic groups (and several phantoms of other, nonexistent, exceptions) and allow for the so-called "twisted forms" which occur when the base field is not algebraically closed, but using the same underlying theories.

The Borovik program, in its various incarnations, accepts that we are studying what is ultimately an algebraic problem involving the pure group language and that there does not appear to be a more purely model theoretic route toward significant structural results. At the same time we have learned a good deal more about what pure model theory has to contribute, notably from the direction of model theory of fields (this is also the case in the study of permutation groups of finite Morley rank).

We will come back to all of this in a more concrete spirit below, in terms of how this works out in the context of the theory described in this paper. (The material of [3] relies for the most part on fundamentally different methods which arose much later, within the specific context of the classification of the finite simple group, and which were in fact still undergoing development as that project came to an end.)

B6. ***Identification theorems.*** More concretely, the developments just mentioned provide the methods used in the study of simple groups of finite Morley rank in odd type to identify the groups in favorable cases.

In high Prüfer 2-rank one can use a method of *Curtis–Tits–Phan* as a way of recovering the Steinberg–Chevalley presentation efficiently, after sufficient structural analysis, and in Prüfer 2-rank 2 one may use the theory of *BN-pairs* of finite Morley rank for the same purpose, the latter generally requiring a more detailed structural analysis. Here, as generally, see [3] for a detailed review of how that actually works.

The model for the treatment in high Prüfer rank is [7][23] and the corresponding axiomatization in Section 4 can be taken as the definition, for our purposes, of the Curtis–Tits–Phan approach.

[23]Or possibly the sharper [8], which retains a K^* hypothesis but in a more limited form. This is published only on arXiv, but finds application in [9].

Our discussion of the case of Prüfer 2-rank 2 (and 2-rank at least 3) in Section 5 does not say much about the final identification or the underlying theory of BN-pairs. Our discussion here has tended to focus on 2-tori as an approximation to algebraic tori. The general thrust of the theory of BN-pairs is to identify a "Borel" subgroup B and the "normalizer N of a maximal torus in B" with properties sufficient to characterize the ambient group.

We work with the group $N = N(T_2)$ and with a certain subgroup $B = TU$ of G, where the construction of U is one of the main difficulties (this is blocked by one bad configuration in the case in which the target group is G_2).

The relevant identification theorem for our purposes is supplied by [48; 49] and once one has identified suitable groups B, N, the structural information required to apply the identification theorem comes down to a verification that the root subgroups of U with respect to T (and their opposites) can be labeled so as to give the expected action of the Weyl group (as well as some loose structural information of the sort given in an explicit form by the Chevalley commutator formula).

In [3] we discussed these two approaches to identification in detail in Sections 6, 7, and 10 of Chapter III, which was devoted to a number of "Specialized Topics" under the broad heading of "Methods". Section 9 of that chapter discusses the signalizer functor theory, which is a considerably more specialized topic that lies more or less at the center of the technical concerns of the present discussion (see above). It is in fact one of the main tools for actually bringing the structural analysis to the point where the standard approaches to identification can be applied.

B7. ***Solvable group theory; Carter subgroups, unipotence theory.*** The point of view taken by L^* theory in odd type makes only limited use of solvable group theory, when compared to the prior K^* theory, which makes very good use of it, notably the parts that go beyond the "basic theory". Of particular importance in that context are the *Borel subgroups*, which as usual are maximal connected solvable subgroups. In the L^* setting one makes good use of the Fitting subgroup and, occasionally, a slightly larger solvable subgroup of momentary interest, but the formal analog of "solvable group" would be "D-group", for which there is not much of a theory in existence, or expected.

From the basic part of the theory comes, in particular, the theory of Hall subgroups (and, in particular, Sylow theory in full generality), the Fitting subgroup, and a good structure theory for nilpotent subgroups. At a more sophisticated level one has *Carter subgroup theory.* Carter subgroups are classically defined in the context of finite solvable groups as self-normalizing nilpotent subgroups, and provide a kind of analog of maximal tori in a general setting. In particular, they are conjugate.

In the context of groups of finite Morley rank, Carter subgroups are taken rather to be definable nilpotent almost-selfnormalizing subgroups (that is, of finite index

Symbol	*Finite version*	*Our version*	*See*
$C_G(X)$	centralizer	(same)	Section B4
$E(H)$	$\left\langle \text{quasisimple subnormal components} \right\rangle$	(same)	Section B4
$F(H)$	Fitting subgroup	(same)	Proposition 3.1
$N_G(H)$	normalizer	(same)	
$O(H)$	odd order radical	*connected* degenerate radical	Fact 4.2
$O^\sigma(H)$	not used	$O(\sigma) = \sigma(O)$	p. 524
$\sigma(H)$	solvable radical	(same)	
U_p	not used	$\left\langle p\text{-unipotent subgroups} \right\rangle$	Lemma 1.6
$U_{(0,r)}$	inconceivable	characteristic 0 unipotence	Section B4
W_T, W_G	not used	Weyl group	Definition 2.4
H°	not used	connected component	[14]
$C_G^\circ, N_G^\circ, \ldots$	not used	connected component of $C_G, N_G, \ldots$	e.g., Theorem 1.18
$\Gamma_{S,2}$	$\langle N_G(A) \mid m_2(A) = 2 \rangle$	similar (definable hull)	Definition 1.13
$\Gamma_V(G)$	$\langle C_G(E) \mid [V : E] = 2 \rangle$	$\langle C_G^\circ(E) \mid [V : E = 2] \rangle$	Definition 2.1
θ	signalizer functors (ad hoc)	signalizer functors (nilpotent)	Section B4

Table 2. Group theoretic operators.

in their normalizers). In this context, one does not require solvability to prove existence, and in the solvable case one is able to recover a fully satisfactory analog of the classical (finite) theory. The theory also provides a possible point of departure for a truly geometrical approach to the classification problem and issues around the algebraicity conjecture, not fully realized, but playing a very significant role in the development of the theory over the last two decades.

At this point it seems appropriate to simply quote a large portion of [3, pp. 108–109], which refers to §I.8 (*Solvable groups*) and more specifically to §I.8.4 (*Carter subgroups*).

> The Carter subgroup was treated first by Wagner in [50], and a full theory given by Frécon in a series of papers beginning with his thesis [31] [cf. [32]]. . . general and extensive. . . . The detailed theory of solvable groups is particularly relevant to the study of minimal simple connected groups, . . . which comes into its own in the treatment of odd type, where problems are often reduced to the minimal simple case and then handled by close analysis there.
>
> The theory of Carter subgroups is very powerful.

Also noteworthy, though not much exploited, is Frécon's work on the conjugacy problem in general, one of the tours de force of the subject [33].

The unipotence theory, on the other hand, continues to play a strong role in the theory and consequently has been discussed above. It tends to come into the picture in connection with Fitting subgroups of not necessarily solvable groups (either as subgroups, or with respect to the action of abelian subgroups on the Fitting subgroup). The results are rather scattered in the literature and this is one of a number of points that would be worth revisiting in a comprehensive text relating to the methods and results of theory of odd type groups.

References

[1] T. Altınel and J. Burdges, "On analogies between algebraic groups and groups of finite Morley rank", *J. Lond. Math. Soc.* (2) **78**:1 (2008), 213–232. MR Zbl

[2] T. Altınel and J. Wiscons, "Actions of Alt(n) on groups of finite Morley rank without involutions", *Proc. Amer. Math. Soc.* **152**:1 (2024), 391–401. MR Zbl

[3] T. Altınel, A. Borovik, and G. Cherlin, *Simple groups of finite Morley rank*, Math. Surveys and Monographs **145**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[4] T. Altınel, J. Burdges, and O. Frécon, "On Weyl groups in minimal simple groups of finite Morley rank", *Israel J. Math.* **197**:1 (2013), 377–407. MR Zbl

[5] T. Altınel, J. Burdges, and O. Frécon, "Structure of Borel subgroups in simple groups of finite Morley rank", *Israel J. Math.* **208**:1 (2015), 101–162. MR Zbl

[6] A. Baudisch, M. Hils, A. Martin-Pizarro, and F. O. Wagner, "Die böse Farbe", *J. Inst. Math. Jussieu* **8**:3 (2009), 415–443. MR Zbl

[7] A. Berkman and A. Borovik, "A generic identification theorem for groups of finite Morley rank", *J. London Math. Soc.* (2) **69**:1 (2004), 14–26. MR Zbl

[8] A. Berkman and A. Borovik, "A generic identification theorem for groups of finite Morley rank, revisited", preprint, 2011. arXiv 1111.6037

[9] A. Berkman and A. Borovik, "Groups of finite Morley rank with a pseudoreflection action", *J. Algebra* **368** (2012), 237–250. MR Zbl

[10] A. Berkman and A. Borovik, "Groups of finite Morley rank with a generically multiply transitive action on an abelian group", *Model Theory* **1**:1 (2022), 1–14. MR Zbl

[11] A. Borovik, "Simple locally finite groups of finite Morley rank and odd type", pp. 247–284 in *Finite and locally finite groups* (Istanbul, 1994), edited by B. Hartley et al., NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci. **471**, Kluwer Acad. Publ., Dordrecht, 1995. MR Zbl

[12] A. Borovik, "A view from lockdown: mathematics discovered, invented, and inherited", pp. 29–34 in *Math in the time of corona*, edited by A. Wonders, Springer, 2021. MR

[13] A. Borovik, "Finite group actions on abelian groups of finite Morley rank", *Model Theory* **3**:2 (2024), 539–569.

[14] A. Borovik and A. Nesin, *Groups of finite Morley rank*, Oxford Logic Guides **26**, Oxford Univ. Press, 1994. MR Zbl

[15] A. Borovik, J. Burdges, and G. Cherlin, "Involutions in groups of finite Morley rank of degenerate type", *Selecta Math.* (*N.S.*) **13**:1 (2007), 1–22. MR Zbl

[16] A. Borovik, J. Burdges, and G. Cherlin, "Simple groups of finite Morley rank of unipotent type", pp. 47–61 in *Algebra*, *logic*, *set theory*, edited by B. Löwe, Stud. Log. (Lond.) **4**, Coll. Publ., London, 2007. MR Zbl

[17] A. Borovik, J. Burdges, and A. Nesin, "Uniqueness cases in odd-type groups of finite Morley rank", *J. Lond. Math. Soc.* (2) **77**:1 (2008), 240–252. MR Zbl

[18] J. Burdges, "Signalizers and balance in groups of finite Morley rank", *J. Algebra* **321**:5 (2009), 1383–1406. MR Zbl

[19] J. Burdges and G. Cherlin, "A generation theorem for groups of finite Morley rank", *J. Math. Log.* **8**:2 (2008), 163–195. MR Zbl

[20] J. Burdges and G. Cherlin, "Semisimple torsion in groups of finite Morley rank", *J. Math. Log.* **9**:2 (2009), 183–200. MR Zbl

[21] J. Burdges and G. Cherlin, "L^*-Groups of odd type with restricted 2-toral automorphisms, I: High Prüfer 2-rank", preprint, 2022, available at https://sites.math.rutgers.edu/∼cherlin/Paper/inprep.html.

[22] J. Burdges and G. Cherlin, "L^*-Groups of odd type with restricted 2-toral automorphisms, II: Prüfer 2-rank 2 and 2-rank at least 3", preprint, 2022, available at https://sites.math.rutgers.edu/∼cherlin/Paper/inprep.html.

[23] J. Burdges and G. Cherlin, "L^*-Groups of odd type with restricted 2-toral automorphisms, III: Identification of PSp_4", preprint, 2022, available at https://sites.math.rutgers.edu/∼cherlin/Paper/inprep.html.

[24] J. Burdges and G. Cherlin, "L^*-Groups of odd type with restricted 2-toral automorphisms, IV: Toward the identification of G_2", preprint, 2022, available at https://sites.math.rutgers.edu/∼cherlin/Paper/inprep.html.

[25] J. Burdges and A. Deloro, "Weyl groups of small groups of finite Morley rank", *Israel J. Math.* **179** (2010), 403–423. MR Zbl

[26] J. Burdges, G. Cherlin, and É. Jaligot, "Minimal connected simple groups of finite Morley rank with strongly embedded subgroups", *J. Algebra* **314**:2 (2007), 581–612. MR Zbl

[27] L. J. Corredor, A. Deloro, and J. Wiscons, "Sym(n)- and Alt(n)-modules with an additive dimension", *J. Algebra* **623** (2023), 1–33. MR Zbl

[28] A. Deloro, "p-Rank and p-groups in algebraic groups", *Turkish J. Math.* **36**:4 (2012), 578–582. MR Zbl

[29] A. Deloro, "Zilber's skew-field lemma", *Model Theory* **3**:2 (2024), 571–586.

[30] A. Deloro and É. Jaligot, "Involutive automorphisms of $N_\circ^\circ$-groups of finite Morley rank", *Pacific J. Math.* **285**:1 (2016), 111–184. MR Zbl

[31] O. Frécon, *Étude des groupes résolubles de rang de Morley fini*, Ph.D. thesis, Université Claude Bernard – Lyon 1, 2000, available at http://www-math.sp2mi.univ-poitiers.fr/∼frecon/These.pdf.

[32] O. Frécon, "Sous-groupes anormaux dans les groupes de rang de Morley fini résolubles", *J. Algebra* **229**:1 (2000), 118–152. MR Zbl

[33] O. Frécon, "Conjugacy of Carter subgroups in groups of finite Morley rank", *J. Math. Log.* **8**:1 (2008), 41–92. MR Zbl

[34] O. Frécon, "Automorphism groups of small simple groups of finite Morley rank", *Proc. Amer. Math. Soc.* **138**:7 (2010), 2591–2599. MR Zbl

[35] J. Freitag and A. Minchenko, "Superstability and central extensions of algebraic groups", *Adv. in Appl. Math.* **72** (2016), 215–230. MR Zbl

[36] J. Freitag, R. Jaoui, and R. Moosa, "The degree of nonminimality is at most 2", *J. Math. Log.* **23**:3 (2023), art. id. 2250031. MR Zbl

[37] W. Hodges, "Meeting Boris Zilber", *Model Theory* **3**:2 (2024), 203–211.

[38] E. Hrushovski, "On pseudo-finite dimensions", *Notre Dame J. Form. Log.* **54**:3-4 (2013), 463–495. MR Zbl

[39] O. H. Kegel and B. A. F. Wehrfritz, *Locally finite groups*, North-Holland Math. Lib. **3**, North-Holland Publishing Co., Amsterdam, 1973. MR Zbl

[40] L. Kramer, K. Tent, and H. Van Maldeghem, "Simple groups of finite Morley rank and Tits buildings", *Israel J. Math.* **109** (1999), 189–224. MR Zbl

[41] A. H. Lachlan, "Two conjectures regarding the stability of ω-categorical theories", *Fund. Math.* **81**:2 (1973/74), 133–145. MR Zbl

[42] A. Macintyre, "On ω_1-categorical theories of fields", *Fund. Math.* **71**:1 (1971), 1–25. MR Zbl

[43] W. E. Marsh, *On ω_1-categorical but not ω-categorical theories*, Ph.D. thesis, Dartmouth College, 1966, available at http://wilfridhodges.co.uk/marsh.pdf.

[44] I. Pak, "The guest publishing scam", blog entry, 2020, available at https://igorpak.wordpress.com/2020/10/26/the-guest-publishing-scam/.

[45] B. Poizat, *Groupes stables*, Nur al-Mantiq wal-Ma'rifa **2**, Bruno Poizat, Villeurbanne, 1987. MR Zbl

[46] B. Poizat, "Quelques mauvais corps de rang infini", pp. 349–365 in *Model theory and applications*, edited by L. Bélair et al., Quad. Mat. **11**, Aracne, Rome, 2002. MR Zbl

[47] B. Poizat, "La conjecture d'algébricité, dans une perspective historique, et surtout modèle-théorique", *Model Theory* **3**:2 (2024), 479–504.

[48] K. Tent, "Split BN-pairs of rank 2: the octagons", *Adv. Math.* **181**:2 (2004), 308–320. MR Zbl

[49] K. Tent and H. Van Maldeghem, "Moufang polygons and irreducible spherical BN-pairs of rank 2, I", *Adv. Math.* **174**:2 (2003), 254–265. MR Zbl

[50] F. O. Wagner, *Stable groups*, London Math. Soc. Lect. Note Ser. **240**, Cambridge Univ. Press, 1997. MR Zbl

[51] F. O. Wagner, "Bad fields in positive characteristic", *Bull. London Math. Soc.* **35**:4 (2003), 499–502. MR Zbl

[52] R. Waldecker, *Isolated involutions in finite groups*, Mem. Amer. Math. Soc. **1061**, Amer. Math. Soc., Providence, RI, 2013. MR Zbl

[53] B. I. Zilber, "The structure of models of categorical theories and the finite-axiomatizability problem", unpublished manuscript, 1977. In Russian.

Received 3 Dec 2022. Revised 28 Feb 2023.

GREGORY CHERLIN:

cherlin.math@gmail.com
Department of Mathematics, Rutgers University, Piscataway, NJ, United States

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.539

Finite group actions on abelian groups of finite Morley rank

Alexandre Borovik

Dedicated to Boris Zilber, who laid the path.

This paper develops some general results about actions of finite groups on infinite abelian groups of exponent p in the finite Morley rank category. These results are applicable to a range of problems on groups of finite Morley rank. Also, they yield a proof of the long-standing conjecture of linearity of irreducible definable actions of simple algebraic groups on elementary abelian p-groups of finite Morley rank. Crucially, these results are needed for the papers by Ayşe Berkman and myself where we have proved an explicit, and best possible, upper bound for the degree of generic multiple transitivity for an action of a group of finite Morley rank on an abelian group.

Preamble

No man is an Iland, intire of it selfe; every man is a peece of the Continent, a part of the maine.
John Donne, 1623

The field of study reflected in the title of this paper could appear to be rather esoteric; however, it is a part of a much wider area of classification of *simple groups of finite Morley rank*. I recommend Gregory Cherlin's informative and incisive survey [26] of the current state of this classification. In short, there are three types of connected simple groups of finite Morley rank: *degenerate* (they do not contain involutions), *odd* (contain involutions, but do not contain infinite groups of exponent 2), and *even* (contain infinite groups of exponent 2). Groups of even type have been identified as simple algebraic groups over algebraically closed fields of characteristic 2 [2]. Little is known about infinite simple groups of degenerate type beyond a fantastic result by Frécon on groups of Morley rank 3 [40] (beautifully elucidated by Corredor and Deloro [29]). On the contrary, quite a lot is known about groups of odd type, but still not enough for proving for them the special case of the *Cherlin–Zilber algebraicity conjecture*:

MSC2020: 03C60, 20F11.
Keywords: groups of finite Morley rank.

> Simple groups of finite Morley rank and odd type are algebraic groups over algebraically closed fields of odd or zero characteristic.

Since the appearance of the classification of simple groups of even type [2], the study of simple groups of finite Morley rank has been moving in two streams:

Stream 1: Proving the Cherlin–Zilber algebraicity conjecture for groups of odd type, as outlined in Cherlin's survey [26].

Stream 2: Study of actions, and specifically actions as automorphisms of abelian groups, of groups of finite Morley rank on the basis of the knowledge already accumulated in the efforts to classify the simple ones. This direction is motivated by the fact that groups of finite Morley rank appeared as *binding groups* in model theory, and hence *act* somewhere. A survey of this direction can be found in [15].

Stream 2 was initiated by Cherlin who suggested to me to start looking for areas of possible application of the classification of groups of even type [2]. After some discussion we decided to take a look at *definably primitive* permutation groups (G, X), that is, definable faithful actions of G on X such that there are no nontrivial definable G-invariant equivalence relations on X. Since the stabilisers of points and orbits of G on X are definable, this means, in particular, that a definably primitive action is transitive. "Faithful" here means that only 1 fixes all elements of X. We proved the following.

Theorem [14, Theorem 1]. *There exists a function $f : \mathbb{N} \to \mathbb{N}$ with the following property. If a group G of finite Morley rank acts on a set X of finite Morley rank definably and definably primitively*, *then*

$$\mathrm{rk}(G) \leqslant f(\mathrm{rk}\, X).$$

The proof of this result is an indicator of the role of the classification technique in Stream 2: an answer to a basic question about actions of groups of finite Morley rank required the use of the classification of simple groups of even type together with the full range of techniques developed for the ongoing study of groups of odd type.

Macpherson and Pillay [49] (following an established tradition from finite group theory) say that a group G of finite Morley rank is of *affine type* if G is a semidirect product of definable subgroups $G = V \rtimes H$, where V is either elementary abelian or divisible torsion-free abelian, and the group H acts on V faithfully and V does not leave any definable subgroup of V other than 0 or V invariant. In that case the natural action of G on the coset space G/H is definably primitive.

This is an important class of primitive groups of permutations. In finite group theory, this class made its first appearance in the celebrated theorem by Galois:

> *A finite solvable primitive permutation group has degree p^k (that is, the set on which it acts contains p^k points) for some prime number p.*

and was the reason why Galois constructed Galois fields [52].

To cut a long story short, the present paper has been written because its results are essential for the study of primitive permutation groups of finite Morley rank and affine type carried out, over some years, by Ayşe Berkman and myself [4; 5; 6]. Our project calls on a surprising range of results and techniques, including almost everything which has been done so far in various approaches to the Cherlin–Zilber conjecture. In particular, the present paper uses basic concepts from the representation theory of finite groups and associative algebras [31; 38] appropriately adapted for the finite Morley rank context.

Finally, a word about the epigraph from John Donne. I proved the results of the present paper in a de facto imprisonment of a strict lockdown.[1] I would not even try to prove them if I did not see my work as part of a much bigger collective project. The lockdown episode reminded me that I started my work in groups of finite Morley rank 40 years ago, also in almost complete isolation, but got critically important help from *our* (as I can now call it) community. This was how I described it in the introduction to [2].

> Vladimir Nikanorovich Remeslennikov in 1982 drew my attention to Gregory Cherlin's paper [24] on groups of finite Morley rank and conjectured that some ideas from my work [on periodic linear groups] could be used in this then new area of algebra. A year later Simon Thomas sent to me the manuscripts of his work on locally finite groups of finite Morley rank. Besides many interesting results and observations his manuscripts contained also an exposition of Boris Zilber's fundamental results on $\aleph_1$-categorical structures which were made known to many Western model theorists in Wilfrid Hodges' translation of Zilber's paper [67] but which, because of the regrettably restricted form of publication of the Russian original, remained unknown to me.

You can learn more about this story from Wilfrid Hodges [43]; a historic perspective is presented by Bruno Poizat [56]. In the early 1980s Hodges worked hard on development of links and channels of communication between Western and Soviet model theorists — and, ironically, also directed me to Zilber's works. I knew Boris personally, but we lived in different cities in Siberia which was as if we were on different planets. In years that followed I have learnt a lot from Boris, but here I wish to emphasise perhaps the most important lesson: the importance of looking at the wider landscapes of mathematics, something that I am trying to do in this paper.

[1] I told this bizarre story in [12].

1. A more technical introduction

1A. ***Ranked universes.*** First of all, we work in a ranked universe in the sense of [16, Chapter 4]. In particular, a *group of finite Morley rank* means for us "*a group definable in a ranked universe*". When we have several algebraic structures of finite Morley rank in the same statement, with definable actions and relations between them, it means that all these structures and relations belong to the same ranked universe $\mathcal{U}$ (which is usually not mentioned). This convention is convenient because it simplifies the language and makes arguments easily accessible to group theorists with some knowledge of the finite group theory or the theory of linear algebraic groups. So far I am aware of just one paper [60] where the expression "a group definable in a ranked universe" is used systematically.

In applications of results of the present paper the universe $\mathcal{U}$ is usually the universe of interpretable sets of some group $\mathcal{G}$ of finite Morley rank. For example, in one of the principal results of the present paper, Theorem 1.5, the group $\mathcal{G}$ is the semidirect product $V \rtimes G$. In its turn, Theorem 1.5 will be applied to a point stabiliser in some generically multiply transitive permutation group $\mathcal{H}$ of finite Morley rank; see further discussion in [6].

1B. ***Some terminology for group actions.*** We use terminology and notation from the books [2; 16] and keep in mind the ranked universe convention of Section 1A.

Let V be an infinite abelian group of finite Morley rank, and X a finite set of definable isomorphisms from V onto V closed under composition and inversion, so X is a finite group. We say in this situation that the finite group X acts on V *definably*. We include the elements of X in the signature of the language and treat them as function symbols. We also say that V is an X-module or that (V, X) is an X-module.

This paper is restricted to the most important case when V is elementary abelian, that is, abelian and periodic of exponent p for some prime number p. We usually treat V as a vector space over the prime field $\mathbb{F}_p$.

We use additive notation for the group operation in V. The key player in our study is the *subring* R generated by X in the ring $\operatorname{End} V$ of endomorphisms of V. Obviously, R is finite and can be viewed as a finite-dimensional $\mathbb{F}_p$-algebra. It is important to observe that elements of R are *definable* endomorphisms of V. We use the usual name for R: it is the *enveloping ring* (or *enveloping algebra over* $\mathbb{F}_p$) of the action of the group X on V and is denoted

$$R = E(X).$$

In a more general situation, if G is any other group which acts on the group V, we say that the action of G is *irreducible* if 0 and V are the only G-invariant subgroups of V; we also say that V is a *simple G-module* or *simple $R(G)$-module*. In our setup, the action of the finite group X on V cannot be irreducible: take $v \neq 0$, then the orbit of v under the action of X is finite and generates a finite X-invariant subgroup, while V is infinite. Therefore we need to adjust the concept of irreducibility to make it usable for the group X.

We say that the action of the group X is *smooth* if any X-invariant *connected* definable subgroup of V equals 0 or V, and, equivalently, that V is a *smooth X-module*. This is a very natural concept, and examples are abundant. For example, let $G = \mathrm{GL}_n(K)$ for an algebraically closed field K of characteristic $p > 0$, acting naturally on $V = K^n$. Then the action of $\mathrm{GL}_n(\mathbb{F}_{p^k}) < G$ on V is smooth.

1C. *Finite groups and Jordan properties.* Our paper starts with a discussion of the following problem, which naturally arises in [5].

Problem 1.1. *Given a finite group X, find good lower bounds for the Morley ranks of faithful smooth X-modules of fixed positive characteristic p.*

The first result of the paper, Theorem 1.2 below, reduces Problem 1.1 to a similar question about the minimal degree of faithful finite-dimensional linear representations of the group X over an algebraically closed field of characteristic p, the latter having been studied in finite group theory for quite some time [36; 46; 61; 62].

Let X be a finite group and p a prime number. We introduce two parameters characterising its size and complexity:

- $d_p(X)$ is the minimal degree of a faithful linear representation of X over the algebraically closed field $\overline{\mathbb{F}}_p$.
- $r_p(X)$ is the minimal Morley rank of an infinite elementary abelian p-group V of finite Morley rank such that X acts on V faithfully, definably, and smoothly.

Theorem 1.2. *Under the assumptions of Problem 1.1*,

$$d_p(X) = r_p(X).$$

This theorem is one of a large body of results which establish close connections and analogies between groups of finite Morley rank, on one hand, and finite groups and algebraic groups, on another.

The following statement is an immediate corollary of Theorem 1.2 via the famous theorem of Larsen and Pink about finite linear groups [47]; see Section 3, Fact 3.1.

Theorem 1.3. *There is a function*

$$J : \mathbb{N} \to \mathbb{N}$$

with the following property:

If H is a finite simple group which acts definably and faithfully on an infinite connected elementary abelian p-group of Morley rank n, then either

- *$|H| \leqslant J(n)$, or*
- *H is a group of Lie type in characteristic p.*

Section 3B contains a brief discussion of Theorem 1.3 and other "theorems of Jordan type"; there is a feeling that there could be some general model-theoretical facts underpinning them all.

1D. ***From finite groups to simple algebraic groups over algebraically closed fields.*** In Sections 4B to 4K we shall study definable actions of simple algebraic groups G over algebraically closed fields on elementary abelian p-groups of finite Morley rank. Our approach is based on the analysis of actions on V of finite subgroups of G, and on the use of the technique developed in Section 2. Theorem 1.4 stated below is the principal tool for transfer of information on certain finite subgroups of G to the group as a whole G itself. The formulation of Theorem 1.4 needs to be preceded by a few words on simple algebraic groups.

First of all, there is some mismatch in the terminology: it is a traditional convention of the theory of algebraic groups that an (infinite) algebraic group G is called *simple* if G is perfect, that is, $[G, G] = G$, with $G/Z(G)$ simple in the usual sense of this word and $Z(G)$ finite. A finite group G with the same properties is called *quasisimple* in finite group theory. So if G is finite, "simple" really means simple: no nontrivial proper normal subgroups. Of course, every finite subgroup is algebraic, but it will be always clear from the context whether a particular algebraic group is finite or not.

Let $G = \mathbb{G}(K)$ be the group of K-points of a simple algebraic group $\mathbb{G}$ defined over an algebraically closed field K of characteristic p. In model theory, it is conventional to call G a simple algebraic group over K. By Poizat [54] the group G is bi-interpretable with the field K (by this we mean that, inside G, field definability implies group definability). On the other hand G is birationally isomorphic with the group of points over K of a simple algebraic group $\mathbb{H}$ defined over the prime field $\mathbb{F}_p$ [8]. Since G and K are bi-interpretable, this isomorphism is definable in G as a pure group; hence we can assume without loss of generality that $\mathbb{G}$ is defined over $\mathbb{F}_p$ and, for every intermediate field $\mathbb{F}_p < F < K$, the group G contains the subgroup $G(F) = \mathbb{G}(F)$ of F-points.

Let now K_∞ be the algebraic closure of the prime field $\mathbb{F}_p$ in K and G_∞ the group of points of G over K_∞. The group G_∞ is the union of finite subgroups $G(\mathbb{F}_{p^k})$ for all natural numbers k, and hence G_∞ is locally finite. As we shall see in later sections, the restriction to the group G_∞ of a definable action of the group G on an elementary abelian p-group of finite Morley rank could be studied by methods developed in Section 2.

Theorem 1.4. *Let K be an algebraically closed field of characteristic $p > 0$ and K_∞ the algebraic closure of the prime field $\mathbb{F}_p$ in K. Let G be a semisimple* *algebraic group over K and G_∞ the group of points of G over K_∞. If M is a subgroup of G containing G_∞ and the structure (G, M) has finite Morley rank, then $M = G$.*

1E. ***Linearisation of actions of simple algebraic groups.*** In Section 4 we prove Theorem 1.5. Here, an action of a group G on an abelian group V is called

- *irreducible* if V contains no G-invariant subgroups other than 0 and V, and
- *definably irreducible* if V contains no G-invariant *definable* subgroups other than 0 and V.

If G is a connected group of finite Morley rank, then these two properties are equivalent [2, Lemma I.11.3].

The following Theorem 1.5 in Section 4C answers the long-standing conjecture of linearity of irreducible definable actions of simple algebraic groups on elementary abelian p-groups of finite Morley rank [16, Question B.38] with further details inquired about in [15, Conjecture 12]. Crucially, Theorem 1.5(1) is needed for the papers by Ayşe Berkman and myself [5; 6]

Theorem 1.5. *Let K be an algebraically closed field of characteristic $p > 0$ and G be a connected algebraic group over K. Assume that G acts definably and faithfully on an elementary abelian p-group V of finite Morley rank. Assume that this action is definably irreducible. Then the following are true*:

(1) *The group V has a structure of a finite-dimensional K-vector space compatible with the action of G.*

(2) *Assume in addition that G is simple. Let $\widehat{G}$ be a simply connected simple algebraic group over K covering G. Then $\rho : \widehat{G} \to G \hookrightarrow \mathrm{GL}(V)$ is an irreducible K-linear representation of the group $\widehat{G}$ on V. There are irreducible* ***rational*** *representations $\omega_1, \ldots, \omega_m$ of the group $\widehat{G}$, and there are $(V \rtimes G)$-definable automorphisms $\varphi_1, \ldots, \varphi_m$ of the field K such that $\rho = \bigotimes_{i=1}^m \varphi_i \omega_i$. In particular, the representation ρ is $(V \rtimes G)$-definable.*

Surprisingly, the following immediate corollary of Theorem 1.5(1) for algebraic groups appears to be new. However, Adrien Deloro informed me that a special case of this result, where G acted transitively on $V \setminus \{0\}$, had been proven in 1983 by Knop [45, Satz 1].

Corollary 1.6. *Let $H = V \rtimes G$ be an algebraic group over an algebraically closed field K of characteristic $p > 0$, where G is a connected algebraic group and V is unipotent (written in additive notation). Assume that V does not have closed G-invariant subgroups other than 0 and V. Then V is an abelian group of exponent p and has a structure of a finite-dimensional vector space over K invariant under the action of G.*

Here, a *unipotent* group in characteristic $p > 0$ is a linear algebraic group containing only p-elements.

Theorem 1.5(2) is a more precise and detailed version of the following result by Bruno Poizat.

Fact 1.7 [55, Theorem 2]. *If K is a field of finite Morley rank and nonzero characteristic p, any simple definable subgroup G of $\mathrm{GL}_n(K)$ is definable in the language of the field K augmented by a finite number of definable field automorphisms.*

In characteristic 0, a much stronger result is known:

Fact 1.8 (a combination of [48; 49; 55], and [13; 35]). *Let (G, V) be a faithful, irreducible module of finite Morley rank, where G is infinite and V is torsion-free. Then there is a definable field over which V is a finite-dimensional vector space and G is a subgroup of $\mathrm{GL}(V)$. If in addition G is simple and contains a nonidentity unipotent element or an involution then G is Zariski closed in $\mathrm{GL}(V)$.*

1F. *Linear groups of finite Morley rank.* Let us return to Theorem 1.5(1) and use notation from its statement. This theorem says that G is a subgroup of the finite-dimensional general linear group $\mathrm{GL}_K(V)$. This is a natural question:

Is the group G Zariski closed in $\mathrm{GL}_K(V)$?

If the automorphisms $\varphi_1, \ldots, \varphi_m$ in part (2) of Theorem 1.5 are Frobenius maps or their inverses, then the representation $G \to \mathrm{GL}_K(V)$ is rational and its image (which is G) is Zariski closed in $\mathrm{GL}_K(V)$. But here we encounter one of the oldest problems of the theory of groups of finite Morley rank.

Problem 1.9 (Angus Macintyre, [16, Question B35, p. 364]). *Can a structure of the form*

$$\langle K; +, \cdot, \varphi \rangle,$$

where $\langle K; +, \cdot \rangle$ is an algebraically closed field of characteristic $p > 0$ and $\varphi \in \mathrm{Aut}(K)$ is neither a Frobenius automorphism nor the inverse of one, have finite Morley rank?

If the answer to Macintyre's problem in *no*, then, in Theorem 1.5, the group G is Zariski closed in $\mathrm{GL}_K(V)$. Otherwise, this is not true in general.

1G. *Linearisation of actions of solvable-by-finite groups.* Finally, we extend Theorem 1.5 to solvable-by-finite groups.

Theorem 1.10. *Let K be an algebraically closed field of characteristic $p > 0$ and G be a connected algebraic group over K. Assume that G acts definably and faithfully on an elementary abelian p-group V of finite Morley rank. Assume that this action is definably irreducible. Then G° is a good torus and V has a definable structure of a finite-dimensional K-vector space compatible with the action of G, with the field K definable in $V \rtimes G$.*

2. Enveloping algebras enter the scene

2A. ***Definitions and generalities.*** In this section we work under assumptions which are weaker than those of Theorem 1.2:

- X is a finite group which acts, definably and smoothly, on an infinite connected elementary abelian p-group V of finite Morley rank.

Notice that we do not assume that the action of the group X is faithful. This allows us to pass these assumptions to factor modules of V by definable X-submodules. The group V is treated as a vector space over $\mathbb{F}_p$.

Lemma 2.1. *The canonical action of the group algebra $A = \mathbb{F}_p[X]$ on V is definable.*

Proof. Indeed every element from A acts on V as a sum of definable endomorphisms (which came from X) and is therefore definable. □

Another important player is the enveloping algebra of X on V, that is, the ring R generated in $\mathrm{End}_{\mathbb{F}_p} V$ by elements of X, or the image of A in $\mathrm{End}_{\mathbb{F}_p} V$ (which is the same). We denote this ring by $E(X)$.

Both A and R are finite-dimensional algebras over $\mathbb{F}_p$, and their action on V is smooth while the action of R on V is also faithful. We treat V as a right A-module and right R-module, and, enlarging the signature of the language, we treat elements from A and R as function symbols.

A finite-dimensional associative algebra over a finite field $\mathbb{F}_p$ of prime order p is the same as a finite ring of characteristic p. Their structure is of course well known. We need a definition of the *Jacobson radical* $J(R)$ of a finite-dimensional algebra R over a field: J is the intersection of all maximal left ideals of R. It can be proved that J is an ideal, and, moreover, J can be characterised as the set of all elements $r \in R$ such that $Mr = 0$ for every simple (or irreducible, which is the same) R-module M.

Fact 2.2 (Wedderburn–Maltsev theorem [37, Theorem VI.2.1]). *Let R be a finite-dimensional associative algebra with identity* 1 *over a finite field $\mathbb{F}_p$ of prime order p and J its Jacobson radical.*

(a) $R = J + S$, *where S is a semisimple algebra*, $J \cap S = 0$, *and S is the direct sum of matrix algebras*

$$S = S_1 \oplus \cdots \oplus S_k, \quad S_i \simeq M_{d_i \times d_i}(\mathbb{F}_{p^{m_i}}),\ i = 1, 2, \ldots, k.$$

(b) *Let $Q = 1 + J$. Then Q is a normal p-subgroup in the group of units R^* of R. Moreover*, *R^* is a semidirect product $R^* = Q \rtimes S^*$ of Q and the group of units S^* of S. In particular*,

$$S^* \simeq \mathrm{GL}_{d_1}(\mathbb{F}_{p^{m_1}}) \times \cdots \times \mathrm{GL}_{d_k}(\mathbb{F}_{p^{m_k}}).$$

2B. ***The case of smooth action.*** The arguments below freely use, without specific references, definitions and results from more basic parts of the theory of modules, which can be found, for example, in [38].

The group X will not be mentioned in the rest of this section. Rather, we work under the following hypothesis.

Hypothesis 2.3. *In the notation of Fact 2.2, R is a finite-dimensional $\mathbb{F}_p$-algebra acting definably, smoothly, and faithfully on an infinite elementary abelian p-group V of finite Morley rank.*

Lemma 2.4. *Under Hypothesis 2.3, the algebra R is semisimple and V is a semisimple R-module.*

Proof. We work in the notation of Fact 2.2. Since VQ is a definable abelian-by-finite p-group, VQ is nilpotent. By properties of commutators in groups of finite Morley rank, $VJ = [V, Q] < V$ is a proper connected definable subgroup of V invariant under R, and hence $VJ = 0$. But then $J = 0$ and R is semisimple; therefore V is also semisimple, that is, a direct sum

$$V = \bigoplus_{\ell \in L} U_\ell \quad \text{for some index set } L \tag{1}$$

of simple R-modules U_ℓ. □

Theorem 2.5. *Under Hypothesis 2.3:*

(a) *All simple factors in the direct sum of equation (1) are isomorphic.*

(b) *The algebra R is simple and therefore is isomorphic to an algebra of all matrices of size $\ell \times \ell$, for some ℓ, over a finite extension of $\mathbb{F}_p$.*

Proof. (a) Denote by $\mathcal{I} = \mathcal{I}(R)$ the set of isomorphism classes of simple R-modules; notice that the set $\mathcal{I}$ is finite. We collect isomorphic simple factors and rewrite the direct sum of simple R-modules in equation (1) as

$$V = \bigoplus_{I \in \mathcal{I}} U_I, \tag{2}$$

where all simple summands of U_I belong to the isomorphism class I.

We want to prove that each submodule U_I is definable; then it follows from the smoothness of the action of R on V than $V = U_I$ for some I, that is, that all summands in (1) are isomorphic.

For that purpose, an element $v \in V$ is called *I-cyclic* for $I \in \mathcal{I}$ if all simple summands in the cyclic module vR belong to I. If u, v are I-cyclic elements of V, then $(u + v)R \leqslant uR + vR$, and all simple summands in $uR + vR$, and hence in $(u + v)R$, belong to I. Therefore $u + v$ is an I-cyclic element. It follows that the set of all I-cyclic elements in V coincides with U_I.

Let us denote by $\mathcal{K}_I$ the set of all right ideals K in R such that all simple summands of the factor module R/K belong to I. Then U_I is defined by the formula

$$\Phi(v) := \bigvee_{K \in \mathcal{K}_I} \Big(\Big(\bigwedge_{k \in K} vk = 0 \Big) \wedge \Big(\bigwedge_{l \in R \setminus K} vl \neq 0 \Big) \Big).$$

This completes the proof of part (a).

(b) Now R is also the enveloping algebra of its restriction to every simple summand of V and is therefore simple. □

2C. ***The weight decomposition for a coprime action of a finite abelian group and the multiplicity formula.*** Now we focus our attention temporarily on actions of finite abelian groups of orders coprime to p, and reformulate the previous results in more familiar terms in this special case.

Let V be a connected H-module of characteristic $p > 0$ with H a finite abelian group of order coprime to p. View V with the action of H as a module over the finite group algebra $A = \mathbb{F}_p[H]$. Applying Maschke's theorem to the action of H on A by multiplication, we see that A is semisimple and is a direct sum of simple finite commutative algebras, that is, finite fields (of course, of characteristic p).

We call a nonzero element $v \in V$ a *weight element* if $\mathrm{Ann}_A(v)$ is a maximal ideal in A; equivalently, this means that vA is an irreducible A-module. It follows that if $F = A/\mathrm{Ann}_A(v)$ and $\lambda : A \to F$ is the canonical homomorphism, then vA is a 1-dimensional vector space over F (notice that F may be bigger than $\mathbb{F}_p$) and, for every $a \in A$, we have

$$va = \lambda(a)v,$$

where the left-hand side is understood in the sense of a (right) A-module, and the right-hand side is a vector space over the field F. This justifies the homomorphisms λ being called *weights* of A.

Observe that when restricted to H, weights become characters of H, that is, homomorphisms from H to the multiplicative group $\overline{\mathbb{F}}_p^*$ of the algebraic closure $\overline{\mathbb{F}}_p$ of the prime field $\mathbb{F}_p$. Also it is easy to see that $F \simeq \mathbb{F}_p[\lambda[H]]$, where

$$\lambda[H] = \{\lambda(h) : h \in H\}.$$

Obviously, if $u, v \in V$ are weight vectors for the same weight λ then either $u + v = 0$ or $u + v$ is a weight vector for λ. Hence all such vectors form a *definable* A-submodule $V_\lambda \leqslant V$, and it follows from Maschke's theorem that

$$V = \bigoplus V_\lambda,$$

where the direct sum is taken over all weights of H on V. It follows that all weight spaces V_λ are connected, and their total number does not exceed $n = \mathrm{rk}\, V$. It is also useful to keep in mind that V_λ is a vector space under the action of the finite field $F_\lambda = \mathbb{F}_p[\lambda[H]]$. Observe further that F_λ is the enveloping algebra for the action of the group H on V_λ.

If we call $\operatorname{rk} V_\lambda$ the *multiplicity* of the weight λ then, as one would expect, we have the following.

Theorem 2.6 (multiplicity formula). *The sum of multiplicities of weights of H on V equals* $\operatorname{rk} V$.

This statement is called a theorem only because of its importance; its proof is obvious.

2D. ***Proof of Theorem 1.2.*** To prove Theorem 1.2, it would suffice to show that, for any definable and faithful smooth X-module V, $\operatorname{rk} V \geqslant \dim_{\overline{\mathbb{F}}_p} W$ for some faithful $\overline{\mathbb{F}}_p[X]$-module W on which X acts faithfully.

So let V be a definable and faithful smooth X-module of smallest possible Morley rank and R the enveloping algebra of this action. By Theorem 2.5, V is a direct sum of isomorphic simple (in particular, finite) R-modules; obviously, X acts on each of them faithfully.

Now we can look at one of these simple R-modules, say U. Assume that $\dim_{\mathbb{F}_p} U = n$. By Theorem 2.5, R is the matrix algebra $\mathrm{Mat}_{m\times m}(\mathbb{F}_{p^l})$, where $n = m \times l$. In particular, we have a definable action of $R^* = \mathrm{GL}_m(\mathbb{F}_{p^l})$ on U, and, since V is a direct sum of isomorphic copies of U, R^* acts definably on V. The maximal torus H of R^* has m different weights on U and therefore on V. From the multiplicity formula (Theorem 2.6) applied to the action of H we have that $\operatorname{rk} V \geqslant m$. But R^*, and hence the (homomorphic) image of X in R^* has a faithful linear representation over $\overline{\mathbb{F}}_p$ of degree m. Hence $\operatorname{rk} V \geqslant d_p(X)$, which completes the proof of Theorem 1.2. $\square$

2E. ***Theorem 1.2: comments.*** It looks as though the proof of Theorem 1.2 does not use all axioms of finite Morley rank (as given in [2, Section I.2.1] and [16, Section 4.1.2]); it would be interesting to find weaker conditions on the module V under which Theorem 1.2 still holds, in a way similar to that of [30].

I expect that the method outlined here gives also new approaches to some of the problems listed in the survey paper by Adrien Deloro and myself [15].

3. Proof of Theorem 1.3 and Jordan properties

3A. ***A Larsen and Pink type theorem.***

Fact 3.1 [47, Theorem 0.2]. *For every n there exists a constant $J'(n)$ depending only on n such that any finite subgroup Γ of GL_n over any field k possesses normal subgroups $\Gamma_3 \leqslant \Gamma_2 \leqslant \Gamma_1$ such that*

(a) $|\Gamma : \Gamma_1| < J'(n)$.

(b) *Either* $\Gamma_1 = \Gamma_2$, *or* $p := \mathrm{char}(k)$ *is positive and* Γ_1/Γ_2 *is a direct product of finite simple groups of Lie type in characteristic* p.

(c) Γ_2/Γ_3 *is abelian of order not divisible by* $\mathrm{char}(k)$.

(d) *Either* $\Gamma_3 = \{1\}$ *or* $p := \mathrm{char}(k)$ *is positive and* Γ_3 *is a* p*-group.*

Theorem 1.2 and Fact 3.1 immediately yield the following result.

Theorem 3.2. *There is a function*

$$J : \mathbb{N} \to \mathbb{N}$$

with the following property:

If X *is a finite group which acts definably and faithfully on an infinite connected elementary abelian* p*-group of Morley rank* n *then* H *possesses normal subgroups* $X_3 \leqslant X_2 \leqslant X_1$ *such that*

(a) $|X : X_1| < J(n)$.

(b) *Either* $X_1 = X_2$, *or* X_1/X_2 *is a direct product of finite simple groups of Lie type in characteristic* p.

(c) X_2/X_3 *is abelian of order not divisible by* p.

(d) X_3 *is a* p*-group.*

Now Theorem 1.3 is a special case of Theorem 3.2. □

3B. ***Theorems of Jordan type.*** Fact 3.1 and Theorem 3.2 can be called *theorems of Jordan type* since they follow the paradigm set by Camille Jordan in his famous theorem of 1878:

Fact 3.3 [44, p. 114]. *There is a function*

$$J : \mathbb{N} \to \mathbb{N}$$

with the following property: *every finite subgroup of* GL_n *over a field of characteristic* 0 *possesses an abelian normal subgroup of index* $\leqslant J(n)$.

Breuillard [20] gave an exposition of Jordan's original proof in the modern terminology; Collins [28] found the optimal explicit bound for $J(n)$.

The introduction to Guld [42] contains an impressive survey of theorems of Jordan type for finite subgroups of groups arising in complex algebraic geometry and in differential geometry. Due to this assumption, [42] contains a more specific and narrower definition of a *Jordan group*:

> A group G is called *Jordan*, *solvably Jordan* or *nilpotently Jordan of class at most* c ($c \in \mathbb{N}$) if there exists a constant $J \in \mathbb{N}$ such that every finite subgroup $X \leqslant G$ has a subgroup $Y \leqslant X$ such that $|X : Y| \leqslant J$ and Y is abelian, solvable or nilpotent of class at most c, respectively.

Fact 3.4 [42, Theorem 2]. *The birational automorphism group of a variety over a field of characteristic* 0 *is nilpotently Jordan of class at most two.*

There are several variations of definitions of Jordan groups, so it could be more useful to speak about all of them as *Jordan properties*. Bandman and Zarhin [3] gave a survey of results on Jordan properties in automorphism groups of some structures of Kähler geometry. Some results on Jordan properties in positive characteristics and further references can be found in [23; 57].

These results, together with many other results quoted in [42], create a feeling that there could be some underlying model-theoretic concepts and results underpinning them all.

4. Linearisation of the actions of algebraic groups

In this section, we prove Theorems 1.4, 1.5, and 1.10.

4A. ***Proof of Theorem 1.4.*** We use the definitions and terminology of Section 1D.

Theorem 1.4. *Let K be an algebraically closed field of characteristic $p > 0$ and K_∞ the algebraic closure of the prime field $\mathbb{F}_p$ in K. Let G be a semisimple algebraic group over K and G_∞ the group of points of G over K_∞. If M is a subgroup of G containing G_∞ and the structure (G, M) has finite Morley rank, then $M = G$.*

Proof. It obviously suffices to consider only the case when G is simple. Let T_∞ be a maximal torus in G_∞. Its Zariski closure in G is a maximal torus in G; let us denote it T. Obviously,

$$T_\infty \leqslant M \cap T \leqslant T,$$

with $M \cap T$ being a definable subgroup. By Poizat [54] the simple algebraic group G and the field K are bi-interpretable. This allows us to apply [2, Proposition I.4.20 and Lemma I.4.21], and prove that T is a *good torus* in the sense of [2, Section I.4.4], that is, every definable subgroup of T is the definable hull of its torsion part. In particular, T is the definable hull of its torsion T_∞. Hence $M \cap T = T$, which means that $T \leqslant M$. Define

$$N = \langle T^m \mid m \in M \rangle.$$

Being generated by Zariski closed connected subgroups, N is Zariski closed. Obviously, $G_\infty = \langle T_\infty^g \mid g \in G_\infty \rangle \leqslant N$. But the Zariski closure of G_∞ in G is G, whence $N = G$ and therefore $M = G$. □

Remark. The bi-interpretability of G and K is the cornerstone of the proof. Indeed, if K is an algebraically closed field and $T = K^*$ is its multiplicative group, then the torus T, *viewed as a pure group*, in absence of the field K, is not a good torus: it is easy to see that the structure (T_∞, T) has Morley rank two.

4B. *Proof of Theorem 1.5.* This is the core of the paper.

Theorem 1.5. *Let K be an algebraically closed field of characteristic $p > 0$ and G be a connected algebraic group over K. Assume that G acts definably on an infinite connected elementary abelian p-group V of finite Morley rank. Assume that this action is definably irreducible. Then the following statements are true.*

(1) *The group V has a structure of a finite-dimensional K-vector space compatible with the action of G, so the group G could be viewed as a subgroup of* $\mathrm{GL}(V)$.

(2) *Let $\widehat{G}$ be a simply connected (quasi-)simple algebraic group over K covering G. Then $\rho : \widehat{G} \to G \hookrightarrow \mathrm{GL}(V)$ is an irreducible K-linear representation of the group $\widehat{G}$ on V. There are irreducible **rational** representations $\omega_1, \ldots, \omega_m$ of the group $\widehat{G}$, and there are $(V \rtimes G)$-definable automorphisms $\varphi_1, \ldots, \varphi_d$ of the field K, such that $\rho = \bigotimes_{i=1}^{d} \varphi_i \omega_i$. In particular, the representation ρ is $(V \rtimes G)$-definable.*

The proof of this theorem will spread over Sections 4C–4I.

It is useful to remember the ranked universe convention of Section 1A.

4C. *Linearisation theorem.* Recall that if G is a group of finite Morley rank acting definably on an abelian group V of finite Morley rank, then the action is called *definably irreducible* if the only G-invariant definable subgroups in V are 0 and V.

Fact 4.1 (linearisation theorem). *Let V be an infinite elementary abelian p-group of finite Morley rank and G an infinite group of finite Morley rank acting on V faithfully, definably, and definably irreducibly. Let D be the ring of all definable endomorphisms of V and $Z = C_D(G)$. Assume that Z is infinite.*

(1) *Z is an algebraically closed field definable in $V \rtimes G$ and the action of Z on V gives V a structure of a finite-dimensional Z-vector space (with a Z-linear action of G).*

(2) *The enveloping algebra (over Z) $R = R(G)$ is the full matrix algebra* $\mathrm{End}_Z(V)$.

(3) *R is definable in $V \rtimes G$.*

Proof. Clause (1) is a result by Macpherson and Pillay [49, Theorem 1.2], with a more complete proof given by Deloro in [32]. It also follows from a more general and very illuminating treatment of linearisation of actions of finite Morley rank given in Deloro's "Zilber's skew field lemma" [33]. Clause (2) follows from basic algebra: an irreducible subgroup $G \leqslant \mathrm{GL}_n(Z)$ contains n^2 matrices linearly independent over Z and forming a basis of the matrix algebra $M_{n\times n}(Z)$. Clause (3) follows from (1). □

4D. ***Groups of units of associative algebras over finite fields: ranks.*** Recall that, for a prime number r, the r-rank $m_r(G)$ of a finite group G is the minimal number of generators in a maximal elementary abelian r-subgroup of G. When $p > 2$ we are interested in the case $r = 2$ because elements of order 2 in $\mathrm{GL}_n(\mathbb{F}_p)$ have eigenvalues ± 1 which, of course, belong to $\mathbb{F}_p$ and therefore have a nice and easy to control behaviour. When $p = 2$, we use $r = 3$, since this still gives us some degree of control. The two clauses (a) and (b) in the following Fact 4.2 correspond to the cases when elements of order 2 are semisimple ($p > 2$) or unipotent ($p = 2$).

Fact 4.2. *Let R be a finite-dimensional associative algebra over a finite field $\mathbb{F}_p$ of prime order p and J its radical. In the notation of Fact 2.2, the following hold*:

(a) *Assume that $p > 2$. Then*

$$m_2(R^*) = d_1 + d_2 + \cdots + d_k.$$

(b) *If $p = 2$ then*

$$m_3(R^*) \leqslant \left\lfloor \frac{d_1}{2} \right\rfloor + \left\lfloor \frac{d_2}{2} \right\rfloor + \cdots + \left\lfloor \frac{d_k}{2} \right\rfloor.$$

Proof. It is easy: for proving (a), perhaps a simple reference to the proof of Theorem 2.6 above would suffice. For (b), a proof follows from an elementary fact from finite group theory: a cyclic group of order 3 has only one faithful irreducible representation over the field $\mathbb{F}_2$, and it is of dimension 2. □

4E. ***Proof of Theorem 1.5 part (1).*** We start with a few general observations. If the connected algebraic group G has a nontrivial unipotent radical $U \neq 1$, then $C_V(U) \neq 0$ is a proper definable G-invariant subgroup of V, which contradicts the assumptions of the theorem. Hence $U = 1$ and G is reductive. Set $Z = Z(G)$. If Z is infinite then the theorem follows from Fact 4.1. So we can assume without loss of generality that G is semisimple.

Let $K_\infty \leqslant K$ be the algebraic closure of the prime field $\mathbb{F}_p$, $G_\infty = \mathbb{G}(K_\infty)$, and $R_\infty = E(G_\infty)$ the enveloping algebra of G_∞.

Lemma 4.3. *R_∞ is the matrix algebra $\mathrm{M}_{d\times d}(K_\infty)$ for some natural number d.*

4F. ***Proof of Lemma 4.3.*** For a subgroup $H \leqslant G_\infty$, we denote by ${}^uH$ the subgroup generated in H by all unipotent elements in H.

We analyse the series of subgroups in G_∞:

$$X_k = {}^uG(\mathbb{F}_{p^{(m+k)!}}), \quad k = 1, 2, \ldots.$$

Obviously, they form a chain

$$X_1 < X_2 < X_3 < \cdots.$$

and

$$\bigcup_{i=1}^{\infty} X_i = G_\infty$$

(since G_∞ is generated by unipotent elements).

It is well known that the groups X_k, $k = 2, 3, 4, \ldots$ (that is, with the exception of a few very small groups), are perfect, $X_k = X_k'$, and therefore

(a) The groups X_k for $k > 1$ have no nontrivial characters $X_k \to \overline{\mathbb{F}}_p^*$.

Let $R_k = E(X_k)$ be the enveloping algebra of X_k in its action on V, then

$$R_1 \leqslant R_2 \leqslant R_3 \leqslant \cdots$$

and

$$R_\infty = \bigcup_{i=1}^{\infty} R_i = E(G_\infty).$$

Denote by J_i the Jacobson radical of R_i, $i = 1, 2, \ldots$. Then R_i/J_i is semisimple and

$$R_i/J_i = M_{i1} \oplus \cdots \oplus M_{ik_i},$$

where M_{ij} are matrix algebras over finite fields F_{ij} of degree d_{ij}. Notice that, in view of claim (a) above, $d_{ij} > 1$ for $i > 1$, since X_i have no nontrivial actions in dimension 1.

For $p > 2$ notice further that d_{ij} coincides with the 2-rank of the corresponding group $\mathrm{GL}_{d_{ij}}(F_{ij})$ of invertible elements (Fact 4.2(a)). Therefore

(b) For $p > 2$, each group R_i^* contains an elementary abelian 2-subgroup of 2-rank

$$d_i = d_{i1} + d_{i2} + \cdots + d_{ik_i}.$$

Again applying Fact 4.2 (a) we see that for $p > 2$ the 2-ranks d_i are bounded by $\mathrm{rk}\, V$ in view of Theorem 2.6; for $p = 2$ we similarly have, from Fact 4.2(b), that $d_i \leqslant 3\, \mathrm{rk}\, V$. Hence

(c) The numerical parameters: $d_i, k_i, d_{i1}, d_{i2}, \ldots, d_{ik_i}$ stabilise starting from some i_* as i grows and remain the same for all $i \geqslant i_*$.

From that index i_* on, embeddings $R_i \leqslant R_j$ for $i < j$ can be much better controlled. Indeed,

(d) After appropriately changing the numeration we have embeddings of rings $M_{il} \leqslant M_{jl}$ for $i_* \leqslant i < j$ where the dimensions d_{il} and d_{jl} of M_{il} and M_{jl}, correspondingly, over their centres, correspondingly, are equal: $d_{il} = d_{jl}$.

This leads to

(e) $J_i \leqslant J_j$ for $i_* \leqslant i < j$.

Indeed, if $J_i \not\leqslant J_j$ then $Q_i \not\leqslant Q_j$ and in one of the factor rings M_{jl} the image $Q = \overline{Q}_i \neq 0$. But $\overline{Q}$ is normalised by invertible elements of M_{il}. Let us focus on groups of units and denote $d_{il} = d_{jl} = m$. Since the groups X_i and X_j are perfect, we see a group $H_i = \mathrm{SL}_m(K_i)$ (which contains X_i) embedded into $H_j = \mathrm{SL}_m(K_j)$ (which contains X_j), where K_i and K_j are two finite fields of characteristic $p > 0$, and H_i is normalising a nontrivial p-subgroup Q in H_j, that is, $H_i \leqslant N_{H_j}(Q)$ and is contained in a proper parabolic subgroup of H_j which is obviously impossible: H_i and H_j have the same Lie rank $m-1$, but proper parabolic subgroups in H_j have smaller Lie rank.

Now we have to take a look at the action of J_i on V. Obviously $VJ_i = [V, Q_i]$. The group $V \rtimes Q_i$ is nilpotent and V is its connected component and therefore, by standard properties of nilpotent groups of finite Morley rank, $[V, Q_i]$ is a connected definable proper subgroup of V. Hence

(f) VJ_i is a connected definable proper subgroup of V.

Now we immediately have

(g) $J_\infty = \bigcup_{i=i_*}^{\infty} J_i$ is a nilpotent ideal of R_∞. Moreover, VJ_∞ is a connected definable proper subgroup of V.

If $W = VJ_\infty \neq 0$ then W is a definable proper connected R_∞-invariant, and hence G_∞-invariant, subgroup of V. But then $N_G(W)$ is a definable subgroup of G, and, of course, contains G_∞; hence, by Theorem 1.4, $N_G(W) = G$, which contradicts irreducibility of G on V. This proves

(h) $J_\infty = 0$

We can now complete the proof of the lemma. By Steps (h) and (d), R_∞ is semisimple and

$$R_\infty = M_1 \oplus \cdots \oplus M_k$$

is the direct sum of matrix algebras of degrees d_j over the field K_∞.

Assume that $k > 1$. then

$$V = VM_1 \oplus \cdots \oplus VM_k,$$

where each submodule VM_j is annihilated by M_l for $l \neq j$, and, moreover,

$$VM_J = \bigcap_{l \neq j} \operatorname*{Ann}_V(M_l).$$

By the chain condition, VM_j is the annihilator of a finite set of matrices, therefore it is definable. Moreover, VM_j is normalised by G_∞, therefore $G_\infty \leqslant N_G(VM_j)$ and the latter is a definable subgroup, which again leads to a contradiction with Theorem 1.4 and irreducibility of G on V. Hence $k = 1$ and $R_\infty = M_{d\times d}(K_\infty)$. □

4G. ***Back to proof of Theorem 1.5 part (1).*** Now we can use Lemma 4.3. Let $Z_\infty \simeq K_\infty^*$ be the centre of R_∞. Take $z \in Z_\infty$, $z \neq 0, 1$. Then

$$z = g_1 + \cdots + g_n$$

for some $g_i \in G_\infty$. Some elements $g \in G$ "commute" with z in the sense that

$$vzg = vgz \quad \text{for all } v \in V. \tag{3}$$

We denote by M the set of such elements in G; it is easy to see that this is a subgroup. And here is the key observation: equation (3) can be written as a first order statement in the group language in the group $V \rtimes G$:

$$vg_1g + \cdots + vg_ng = vgg_1 + \cdots + vgg_n \quad \text{for all } v \in V.$$

Now the subgroup M is definable in $V \rtimes G$, and contains G_∞, so we have the following chain of subgroups:

$$G_\infty \leqslant M \leqslant G. \tag{4}$$

The group G, being a semisimple linear algebraic over K, is decomposed as a central product

$$G = G^1 * \cdots * G^\ell$$

of simple algebraic groups G^i over K, $i = 1, \ldots, \ell$. If G^i_∞ is the group of points of G^i over the field K_∞, then

$$G^i_\infty = G^1_\infty * \cdots * G^\ell_\infty,$$

$M_i = M \cap G^i$, and in every G^i we have a chain of subgroups

$$G^i_\infty \leqslant M_i \leqslant G^i.$$

Now, Theorem 1.4 gives us $M_i = G^i$ for all i, so $M = G$. Recall that the subgroup M was constructed from some element $z \in Z$. This argument applies to all $z \in Z_\infty$, so Z_∞ and G commute elementwise as multiplicative subgroups of $\operatorname{End} V$. The application of Fact 4.1 completes the proof of Theorem 1.5(1). □

4H. ***Theorem 1.5 part (2): some preparatory comments.*** We continue to work in the notation of Theorem 1.5 but need additional definitions and facts about simple algebraic groups over algebraically closed fields.

If G and H are two simple algebraic groups over an algebraically closed field K of characteristic $p > 0$ which is fixed in this section, a surjective rational homomorphism $\zeta : G \to H$ is called an *isogeny*. It is known that $\ker \zeta \leqslant Z(G)$ is finite and has order coprime to p. Among simple algebraic groups of the same type as G, that is, with the same root system, there is the group $\widehat{G}$, called *simply* *connected*, which has a isogeny onto any other simple group of the same type, and the *adjoint* group $\overline{G}$ such that any group of the same type has an isogeny onto $\overline{G}$.

Let Φ be the root system of G; we can select root subgroups

$$X_r = \{x_r(t) : t \in K\}, \quad r \in \Phi,$$

so that all of them are defined over the prime field. An isogeny $\zeta : G \to H$ maps the system of root subgroups $\{X_r : r \in \Phi\}$ to a similar system in H.

If now $\varphi \in \operatorname{Aut} K$ is an automorphism of the field K then it induces a *field automorphism* of G by mapping elements of root subgroups of G,

$$x_r(t) \mapsto x_r(t^{\varphi}), \quad r \in \Phi,\ t \in K.$$

This gives us an automorphism $\widetilde{\varphi}$ of G, and similarly for H. Obviously, the $\widetilde{\varphi}$ commute with the isogeny, $\widetilde{\varphi}\zeta = \zeta\widetilde{\varphi}$, which justifies the use of the notation $\widetilde{\varphi}$ on both groups G and H. Also, the action of $\widetilde{\varphi}$ on the elements of the root subgroup $X_r = \{x_r(t) : t \in K\}$ is the same as the action of φ on the elements of K. Therefore, if $\widetilde{\varphi}$ is a definable automorphism of G, then φ is a definable automorphism of K.

The automorphism $\widetilde{\varphi}$ can be used to convert each G-module W into another G-module, denoted W^{φ}, by the rule $wg := w(g^{\widetilde{\varphi}})$, $w \in W$, $g \in G$ [58, §5].

4I. *Proof of Theorem 1.5 part (2).*

Proof. We know from part (1) of Theorem 1.5 that V is a finite-dimensional vector space over the field K and G is a definable subgroup in $\mathrm{GL}_K(V)$.

Let $\widehat{G}$ be a simply connected simple algebraic group over K and $\rho : \widehat{G} \to G$ an isogeny. Then ρ is an *abstract homomorphism*, in the sense of the famous *Homomorphismes "abstraits"* paper by Borel and Tits [10], from $\widehat{G}$ to $\mathrm{GL}_K(V)$,

$$\rho : \widehat{G} \to \mathrm{GL}(V).$$

By [10, Corollary 10.4] (compare with [58]), ρ is equivalent to a tensor product

$$\varphi_1\omega_1 \otimes \cdots \otimes \varphi_m\omega_m,$$

where ω_i are rational irreducible representations of $\widehat{G}$ and φ_i are automorphisms of $\widehat{G}$ induced by automorphisms of the field K.

Now all that we have to prove is that all field automorphisms φ_i are definable in $V \rtimes G$. For that, we have to switch from the representation-theoretic language to the group-theoretic one.

Let ϵ be the 1-dimensional trivial representation $\widehat{G} \to \mathrm{GL}_1(K)$. Set $d_i = \dim_K \omega_i$; by properties of tensor products, $d_1 \cdots d_m = n$, where $n = \dim_K(V)$. As usual, define $d_i\epsilon = \epsilon \oplus \cdots \oplus \epsilon$ (d_i times). Finally, set

$$\rho_i = d_1\epsilon \otimes \cdots \otimes d_{i-1}\epsilon \otimes \varphi_i\omega_i \otimes d_{i+1}\epsilon \otimes \cdots \otimes d_m\epsilon.$$

The representations ρ_i are defined up to equivalence of representations, but they can be replaced by equivalent ones in a way that makes the images G_i of ρ_i pairwise commuting, $[G_i, G_j] = 1$ for $i \neq j$. Set $\breve{G} = G_1 \cdots G_m$. Observe that $G \leqslant \breve{G}$. In particular, the group $\breve{G}$ acts on V irreducibly. Also, the centres $Z(G_i) < Z(\mathrm{GL}(V))$ consist of scalar matrices. Let $\zeta : \mathrm{GL}(V) \to \mathrm{PGL}(V)$ be the canonical homomorphism; replacing all our groups by their images in $\mathrm{PGL}(V)$ simplifies the arguments. The steps of this reduction are shown in diagram (5)

$$\begin{tikzcd}
\widehat{G} \arrow[dd, "\rho=\bigotimes\widetilde{\varphi}_i\omega_i"'] \arrow[ddr, "\rho_i"] & & \text{simply connected cover of } G \\
& & \text{representations} \\
G \arrow[dd, "\zeta"'] & G_i \arrow[dd, "\zeta"'] & \mathrm{GL}(V) \\
& & \mathrm{GL}(V) \to \mathrm{PGL}(V) \\
\overline{G} \arrow[r, "\widetilde{\varphi}_i\overline{\omega}_i"] \arrow[ddr, "\pi_i"] & \overline{G}_i \arrow[dd, "\mathrm{Id}"'] & \mathrm{PGL}(V) \\
& & \text{projections } \overline{G} \to \overline{D} \\
& \overline{G}_i & \mathrm{PGL}(V)
\end{tikzcd} \tag{5}$$

and explained in detail below.

We are moving now into the setup of [10, Theorem 10.3] and denote by $\bar{\rho}$ and $\bar{\rho}_i$ the induced homomorphisms

$$\bar{\rho} = \rho \cdot \zeta : \widehat{G} \to \mathrm{PGL}(V), \qquad \bar{\rho}_i = \rho_i \cdot \zeta : \widehat{G} \to \mathrm{PGL}(V).$$

We denote by $\overline{G}$ and $\overline{G}_i$ the images of groups G and G_i in $\overline{P} = \mathrm{PGL}(V)$. Since $\overline{G}_i = G_i/Z(G_i)$ are simple groups, the commuting product $\overline{G}_1 \cdots \overline{G}_m$ is a direct product,

$$\overline{G}_1 \cdots \overline{G}_m = \overline{G}_1 \times \cdots \times \overline{G}_m.$$

Now we take the double centraliser closures of groups $\overline{G}_i$, setting $\overline{D}_i = C_{\overline{P}}(C_{\overline{P}}(\overline{G}_i))$, then $\overline{G}_i \leqslant \overline{D}_i$ and the groups $\overline{D}_i$ form a direct product

$$\overline{D} = \overline{D}_1 \times \cdots \times \overline{D}_m.$$

It could be shown that $\overline{D}_i \simeq \mathrm{PGL}_{d_i}(K)$, but we will not be using this fact. What matters for us is that $\overline{D}_i$ are definable in $\mathrm{PGL}(V)$ and hence in $V \rtimes G$.

The final step in the proof starts with an observation that $\overline{G} \leqslant \overline{D}$. The group $\overline{G}$ is of course definable, and hence the projection maps $\pi_i : \overline{G} \to \overline{D}_i$ are definable in $V \rtimes G$. The image of π_i is $\overline{G}_i$, and the triangle at the bottom of the diagram (5) is commutative, $\pi_i = \widetilde{\varphi}_i \overline{\omega}_i$ and $\widetilde{\varphi}_i = \pi_i \overline{\omega}_i^{-1}$. Hence the field automorphism $\widetilde{\varphi}_i$ of the group $\overline{G}$ is definable in $V \rtimes G$, and hence the automorphism φ of the field K is also definable in $V \rtimes G$. $\square$

4J. ***Theorem 1.5 part (2): an example.*** Let $\widehat{G} = \mathrm{SL}_2(K)$, where K is an algebraically closed field of characteristic $p > 2$, ω the canonical 2-dimensional representation of $\widehat{G}$ over K and $\widetilde{\varphi}$ its field automorphism induced by $\varphi \in \mathrm{Aut}\, K$. Let V be the space of the representation $\omega \otimes \widetilde{\varphi}\omega$; then $\dim_K V = 4$. Using the usual notation I_n for the identity linear transformation of K^n, we see that the image in $\mathrm{GL}(V)$ of the central element $-I_2$ of $\widehat{G}$ is

$$-I_2 \otimes -I_2^{\widetilde{\varphi}} = -I_2 \otimes -I_2 = I_4,$$

which means that the image of $\widehat{G}$ in $\mathrm{GL}(V)$ (we denote it G) is isomorphic to $\mathrm{PSL}_2(K)$. If we now move to $\mathrm{PGL}(V)$, retaining the notation from the proof, we see that $\overline{G}$ now is a subgroup in $\mathrm{PSL}_2(K) \times \mathrm{PSL}_2(K)$ and that it happens to be exactly the graph of the field automorphism $\widetilde{\varphi} : \mathrm{PSL}_2(K) \to \mathrm{PSL}_2(K)$.

If we now look only at the group $V \rtimes G \simeq K^4 \rtimes \mathrm{PSL}_2(K)$, it would be difficult to find its representation-theoretic origins without invoking the simply connected cover $\widehat{G}$ of the group G. Therefore, Theorem 1.5(2) corrects (and confirms, in the corrected form) [15, Conjecture 9.].

The proof of the theorem in this special case contains an interesting little detail which sheds light at the situation in general: we are proving the definability of the automorphism $\widetilde{\varphi}$ by proving the definability of its graph as a subgroup. This is a cute tiny self-evident fact from elementary algebra which Şükrü Yalçınkaya and myself could not find in any textbook but which we systematically use in all our work on black box algebra [17]:

> *A map $\varphi : G \to H$ from a group G to a group H is a homomorphism if and only if its graph $\Gamma_\varphi \subset G \times H$ of φ is a subgroup of $G \times H$.*

The same is of course true for rings and all other kinds of algebraic systems. This is one of many examples of the exchange of ideas between the theory of black box groups and the theory of groups of finite Morley rank. See also Section 5E for further discussion.

4K. ***Proof of Theorem 1.10.*** The following result about solvable groups of finite Morley rank is an adaptation of the method of proof of Theorem 1.5(1).

Theorem 1.10. *Let G be an infinite solvable-by-finite group of finite Morley rank which acts faithfully and definably on a connected elementary abelian p-group V*

of finite Morley rank. Assume that this action is definably irreducible. Then G° is a good torus and V has a definable structure of a finite-dimensional K-vector space compatible with the action of G, with the field K definable in $V \rtimes G$.

Proof. Since G acts faithfully and irreducibly on V, $[G^\circ, G^\circ]$ acts trivially on V by [2, Lemma I.8.2], and since the action is faithful, $[G^\circ, G^\circ] = 1$ and G° is abelian. By a similar argument, G° is a $p^\perp$-group. By [2, Proposition I.11.7], G° is a good torus. By [2, Fact I.9.5], there is a subgroup $G_\infty < G$ such that $G^\circ \cap G_\infty$ is the torsion part of G° and $G^\circ G_\infty = G$. Obviously, G_∞ is a locally finite group and G is the definable closure of G_∞.

Now we can repeat, with very small changes, the proof of Theorem 1.5(1). □

5. Historical and other comments

I wish to conclude the paper with a few words about the balance of the model-theoretic and the group-theoretic components in the theory of simple groups of finite Morley rank.

5A. ***Simple algebraic groups, Chevalley groups and the work version of the Cherlin–Zilber conjecture.*** In the classification theory of simple groups of finite Morley rank, as it stands now, the structural theory of simple algebraic groups is heavily used, as a rule, in the form of a summary statement:

> *A simple algebraic group over an algebraically closed field K is a Chevalley group over K.*

A Chevalley group over a field or a ring R is viewed as a group given by some specific generators and relations which involve parameters from R [8]. It can be shown that

> *A Chevalley group over an algebraically closed field K is a simple algebraic group over K.*

In works aimed at proving the Cherlin–Zilber algebraicity conjecture, this is usually used in the following form:

> *An infinite simple group of finite Morley rank is a Chevalley group over an algebraically closed field.*

This allows us to use very powerful group-theoretic characterisations of Chevalley groups and ignore the algebraic geometry aspects of the theory. See, for example, how the Curtis–Tits–Phan–Lyons theorem [41], which describes Chevalley groups as amalgams (in the group-theoretic sense) of groups of type SL_2 or PSL_2, is used in [7].

Let G be a simple algebraic group over an algebraically closed field K. Let T be a maximal torus in G. The canonical approach to describing G as a Chevalley group is to associate with T the Weyl group, a finite system Φ of roots, root subgroups, etc. Almost all information about G needed for using G within an attempted proof of the Cherlin–Zilber conjecture is contained in the set of the so-called *root* SL_2-*subgroups*, which can be characterised as Zariski closed subgroups in G isomorphic to $\mathrm{SL}_2(K)$ or $\mathrm{PSL}_2(K)$ and normalised by T. They are labelled by pairs of roots $\pm r \in \Phi$, and they generate G. The origins of the concept go to the paper by Borel and de Siebenthal of 1949 on the structure of compact Lie groups [9]; see the construction of root SL_2-subgroups in [8, 3.2(1)]. The following special case of the Curtis–Tits–Phan–Lyons theorem is formulated in [7, Proposition 2.1.].

Fact 5.1. *Let Φ be an irreducible finite root system of rank at least* 3, *and let Π be a system of fundamental roots for Φ. Let X be a group generated by subgroups X_r for $r \in \Pi$. Suppose that either $[X_r, X_s] = 1$ or $X_{rs} = \langle X_r, X_s\rangle$ is a Chevalley group over an algebraically closed field with the root system Φ_{rs} spanned by r and s, and with X_r and X_s corresponding root* SL_2*-subgroups with respect to some maximal torus of X_{rs}. Then $X/Z(X)$ is isomorphic to a Chevalley group with the root system Φ via a map carrying the subgroups X_r to root* SL_2*-subgroups.*

5B. ***Central extensions of simple algebraic groups.*** However, there is a model-theoretic twist again: all these arguments rely on the description of central extensions of Chevalley groups in the finite Morley rank context due to Tuna Altınel and Gregory Cherlin [1], which, in its turn, relies on model-theoretic results by Newelski and Wagner (independently):

Fact 5.2 ([53; 64], cf. [2, Lemma I.4.16(2)]). *Let K be a field of finite Morley rank and X a definable subgroup of $K^\times$ which contains the multiplicative group of an infinite subfield F of K (not assumed definable). Then $X = K^\times$.*

In particular, [1] allows to conclude that if the subgroup X in Fact 5.1 is of finite Morley rank, then not only $X/Z(X)$, but X itself is a Chevalley group.

5C. ***Good tori.*** It is worth noting that a key ingredient of the proofs of Theorem 1.4 and of Theorem 1.10, a "*good torus*", a concept introduced by Gregory Cherlin [25], is rooted in the model theory. In the book [2], Proposition I.11.7, quoted in the proofs, goes back to Proposition I.4.15, which is a deep model-theoretic result of 2001 by Frank Wagner [64].

5D. ***Bi-interpretability.*** Another key ingredient, the bi-interpretability of the simple algebraic group G over an algebraically closed field K, and the field K, is Bruno Poizat's result [54] of 1988. Its proof is purely model-theoretical and does not use the structural theory or classification of simple algebraic groups. What is even more remarkable, Bruno Poizat does not even assume that G is linear — in his paper, the structure of G as an algebraic variety is proved to be definable in the group language of G. Its prehistory is also interesting: the model-theoretic ideas underpinning the proof can be traced back to Zilber's result of 1977. As Gregory Cherlin discussed it in 1979,

> *[...] Zilber* [66] *gives an elegant proof that a simple algebraic group over an algebraically closed field is $\aleph_1$-categorical [16, Corollary to Theorem 3.2]. I had observed (Fall 1976) that this result can be obtained easily, but at some length, from the known structure theory for such groups (using a good deal of* [27] *and the generators and relations of Steinberg* [59]*). Zilber's proof is short and uses no structure theory.*
> ([24, p. 2], reference numbers are updated to match the bibliography of this paper.)

Not surprisingly, relations between model-theoretic properties between a Chevalley group and its field (not necessary algebraically closed) or a ring of definition are of natural interest.

At the present time, the most powerful result belongs to Elena Bunina [21]:

> *If $G(R) = G_\pi(\Phi, R)$ is a Chevalley group of rank > 1, R is a local ring (with $\frac{1}{2}$ for the root systems $\mathrm{A}_2, \mathrm{B}_l, \mathrm{C}_l, \mathrm{F}_4, \mathrm{G}_2$ and with $\frac{1}{3}$ for G_2), then the group $G(R)$ is regularly bi-interpretable with the ring R.*

As usual, the Chevalley group $G_\pi(\Phi, R)$ is constructed from the root system Φ, a ring R and a representation π of the corresponding Lie algebra [8].

But the bi-interpretability of a Chevalley group over an algebraically closed field K with this field is a relatively easy result.

5E. ***Bi-interpretability in the black box algebra.*** Anatoly Maltsev was the pioneer, in 1961, of the study of bi-interpretability of Chevalley groups and their fields of definition. Theorem 4 of his paper [50] states the bi-interpretability of linear groups $G = \mathrm{GL}_n(K)$, $\mathrm{PGL}_n(K)$, $\mathrm{SL}_{n+1}(K)$, and $\mathrm{PSL}_{n+1}(K)$, $n \geqslant 2$ over a field K and the field K.

Moreover, Maltsev had shown that this bi-interpretability is recursive: there are algorithms which rewrite formulae from $\mathrm{Th}(G)$ as formulae from $\mathrm{Th}(K)$, and vice versa. This algorithmic aspect has interesting and somewhat bizarre analogues in *Black Box Algebra* as developed by Şükrü Yalçınkaya and myself [18]. The black box algebra deals with finite algebraic structures (in particular, Chevalley groups over finite fields), where, however, algebraic operations are computed by "black boxes" (these are finite analogues of definable structures from model theory). Homomorphisms have to be computed (and first order formulae evaluated) by Monte-Carlo algorithms in probabilistic polynomial time. Polynomial time morphisms are analogues of definable homomorphisms from model theory. Interestingly, this approach gives some indication why Maltsev skipped, in his theorem, the groups

$\mathrm{SL}_2(K)$, and $\mathrm{PSL}_2(K)$: in the black box context, we do not have direct access to nontrivial unipotent elements even in these "small" groups. In model theory, their existence is a basic statement

$$(\exists u)(u^p = 1 \wedge u \neq 1), \tag{6}$$

but in the black box groups the quantifier $\exists$ means "can be found in probabilistic polynomial time" and proof of (6) for $\mathrm{SL}_2(K)$ and $\mathrm{PSL}_2(K)$ was believed to be an intractable problem. Even this innocent looking problem was seen as impossibly difficult:

> Assume that you are given several matrices $M_1, \ldots, M_m$ of size $n \times n$ over a finite field $\mathbb{F}_{p^k}$ generating a subgroup X isomorphic to $\mathrm{SL}_2(\mathbb{F}_{p^\ell})$. Find in X a nontrivial unipotent element (that is, an elements of order p).

The reason for that is simple: the probability to hit a unipotent element at random is about $1/p^k$ and is exponentially small with the growth of k, even with the small values of p.

Şükrü Yalçınkaya and myself clarified all that in [17], by constructing $\mathrm{PGL}_2(K)$ from $\mathrm{PSL}_2(K)$ viewed as a pure group, then interpreting a projective plane $\mathbb{P}^2(K)$ in $\mathrm{PGL}_2(K)$, and, finally, interpreting K in $\mathbb{P}^2(K)$ (the last step is well known but has some twists in the polynomial time setting). In bigger Chevalley groups, finding unipotent elements is done by recursion to these "small" cases (see Yalçınkaya [65] and our forthcoming monograph [18]). This was considerably more difficult than Maltsev's analysis in [50] — but still, Maltsev was the pioneer.

Another example of exchange of ideas between the black box group theory and the theory of groups of finite Morley rank was given in Section 4J. In the black box group theory, a group homomorphism $\varphi : G \to H$ with a black box for its graph $\Gamma_\varphi < G \rtimes H$ is not necessarily polynomial time computable; we call it a *protomorphism*. The concept of protomorphism is central to the theory of black box groups.

5F. ***Alternative versions of the proof of Theorem 1.4.*** I will now outline two alternative versions of the proof of Theorem 1.4, which use the structural theory of simple algebraic groups to reduce the proof to the case of simple algebraic groups G of type A_1, that is, $G = \mathrm{PSL}_2(K)$ or $\mathrm{SL}_2(K)$.

Lemma 5.3. *It suffices to prove Theorem 1.4 for G if G is of type A_1, that is, $G = \mathrm{PSL}_2(K)$ or $\mathrm{SL}_2(K)$.*

Proof. Assume that we are in the setup of Theorem 1.4 and that we already know that Theorem 1.4 is true in the special case of G being of type A_1, that is, $G = \mathrm{PSL}_2(K)$ or $\mathrm{SL}_2(K)$.

Let us pick in G_∞ a maximal torus T_∞ and set $T = C_G(T_\infty)$; this is a maximal torus in G, and it equals the Zariski closure of T_∞. Denote by $\mathcal{L}_\infty$ the set of all root SL_2-subgroups in G_∞ normalised by the torus T_∞ and by $\mathcal{L}$ the set of their Zariski closures. Then $\mathcal{L}$ is the set of all root SL_2-subgroups in G normalised by the torus T.

Let now $L \in \mathcal{L}$. Then $L_\infty = L \cap G_\infty$ is (P)$\mathrm{SL}_2(K_\infty)$ and

$$L_\infty \leqslant M \cap L \leqslant L,$$

and by the assumptions of the lemma, $L = M \cap L < M$. Since the system $\mathcal{L}$ generates the group G, $M = G$. This proves the lemma. □

Lemma 5.4. *Theorem 1.4 is true if G is of type A_1, that is, $G = \mathrm{PSL}_2(K)$ or $\mathrm{SL}_2(K)$.*

Proof. There are at least five approaches to a proof of this lemma.

(1) This is the most direct and straightforward approach to the proof: let T_∞ be a torus in G_∞ and $T = C_G(_\infty)$ a torus in G. By basic linear algebra, T is contained in a Borel subgroup of G and is therefore a good torus by [2, Proposition I.11.7]. Hence $M \cap T = T$ and $T < M$. If U_∞ and V_∞ are the two maximal unipotent subgroups in G_∞ normalised by T_∞ then $U = [U_\infty, T]$ and $V = [V_\infty, T]$ are two different maximal unipotent subgroups in G and belong to M. Since $\langle U, V\rangle = G$, we have $M = G$. □

Other four approaches are only indicated:

(2) The lemma immediately follows from Theorem 4 of Bruno Poizat's seminal paper [55] (which, in its turn, is based on the model-theoretic, by their nature, results by Frank Wagner [63; 64]).

(3) Theorem 4 of [55] has been drastically improved by the very neat result of the paper of Mustafin and Poizat [51], which does not use Wagner's theorem: a superstable nonsolvable subgroup of $\mathrm{SL}_2(K)$ is conjugated to $\mathrm{SL}_2(k)$, where k is an algebraically closed subfield of K; it is therefore transparent that if you assume, in addition, that the pair of groups has a finite Morley rank, so has the pair of fields, and $K = k$.

(4) In odd characteristics, the lemma is an almost immediate consequence of the difficult and important result by Adrien Deloro and the late Éric Jaligot [34]. Indeed it follows from the well-known properties of the group $\mathrm{PSL}_2(K)$ that the subgroup M satisfies the assumptions of their theorem, and therefore M is isomorphic to $\mathrm{PSL}_2(F)$ for some algebraically closed field F (of the same characteristic as K, of course); after that it becomes obvious that $M = G$.

(5) In characteristic 2, the lemma follows from [19]. Moreover, it de facto follows from an ancient result (the unbelievable 1900!) by Burnside [22]; see discussion in [11, Sections 4 and 5] and in [39, pp. 11–12].

To my taste, approach (1) is the simplest and best fits the needs of classification of simple groups of finite Morley rank. □

Acknowledgements

My special thanks go to Ayşe Berkman, my research collaborator and coauthor of many years. Problems solved in this paper originate in our joint project [4; 5; 6].

I was much influenced by ideas that my coauthors Tuna Altınel, Ayşe Berkman, Jeff Burdges, Adrien Deloro, Gregory Cherlin, Ali Nesin, and Şükrü Yalçınkaya shared with me over years of our collaboration.

Gregory Cherlin kindly copy-edited an earlier version of this paper; any grammatical and orthographic errors introduced by me later are my own responsibility. Yuri Zarhin sent to me his paper [3], which extended the range of Jordan subgroups. Theorem 1.5 and Corollary 1.6 were first proved for simple groups G; Gregory Cherlin suggested that the result remains valid for connected groups. Questions, comments and advice from Adrien Deloro, Gregory Cherlin, Bruno Poizat, and Şükrü Yalçınkaya significantly improved the paper and led to new results — for example, to Theorem 1.5 part (2).

The anonymous referee made three consecutive rounds of comments which really helped to improve the text.

To summarise, I have had *community support* to my work — and this why I quote John Donne in the Preamble to this paper.

References

[1] T. Altınel and G. Cherlin, "On central extensions of algebraic groups", *J. Symbolic Logic* **64**:1 (1999), 68–74. MR Zbl

[2] T. Altınel, A. V. Borovik, and G. Cherlin, *Simple groups of finite Morley rank*, Math. Surveys and Monographs **145**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[3] T. Bandman and Y. G. Zarhin, "Automorphism groups of $\mathbb{P}^1$-bundles over a non-uniruled base", *Uspekhi Mat. Nauk* **78**:1 (2023), 3–66. In Russian; translated in *Russian Math. Surveys* **78**:1 (2023), 1–64. MR Zbl

[4] A. Berkman and A. Borovik, "Groups of finite Morley rank with a generically sharply multiply transitive action", *J. Algebra* **513** (2018), 113–132. MR Zbl

[5] A. Berkman and A. Borovik, "Groups of finite Morley rank with a generically multiply transitive action on an abelian group", *Model Theory* **1**:1 (2022), 1–14. MR Zbl

[6] A. Berkman and A. Borovik, "Groups acting generically multiply transitively on solvable groups", preprint, 2024. arXiv 2402.06987

[7] A. Berkman, A. V. Borovik, J. Burdges, and G. Cherlin, "A generic identification theorem for L^*-groups of finite Morley rank", *J. Algebra* **319**:1 (2008), 50–76. MR Zbl

[8] A. Borel, "Properties and linear representations of Chevalley groups", pp. 1–55 in *Seminar on algebraic groups and related finite groups* (Princeton, NJ, 1968/69), Lecture Notes in Math. **131**, Springer, 1970. MR Zbl

[9] A. Borel and J. De Siebenthal, "Les sous-groupes fermés de rang maximum des groupes de Lie clos", *Comment. Math. Helv.* **23** (1949), 200–221. MR Zbl

[10] A. Borel and J. Tits, "Homomorphismes "abstraits" de groupes algébriques simples", *Ann. of Math.* (2) **97** (1973), 499–571. MR Zbl

[11] A. Borovik, "Mathematics discovered, invented, and inherited", preprint, 2013. arXiv 1309.3073

[12] A. Borovik, "A view from lockdown: mathematics discovered, invented, and inherited", pp. 29–34 in *Math in the time of corona*, edited by A. Wonders, Springer, 2021. MR

[13] A. Borovik and J. Burdges, "Linear groups of finite Morley rank", preprint, 2008. arXiv 0801.3958

[14] A. Borovik and G. Cherlin, "Permutation groups of finite Morley rank", pp. 59–124 in *Model theory with applications to algebra and analysis*, vol. 2, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **350**, Cambridge Univ. Press, 2008. MR Zbl

[15] A. Borovik and A. Deloro, "Binding groups, permutation groups and modules of finite Morley rank", preprint, 2019. arXiv 1909.02813

[16] A. Borovik and A. Nesin, *Groups of finite Morley rank*, Oxford Logic Guides **26**, Oxford Univ. Press, 1994. MR Zbl

[17] A. Borovik and Ş. Yalçınkaya, "Adjoint representations of black box groups $\mathrm{PSL}_2(\mathbb{F}_q)$", *J. Algebra* **506** (2018), 540–591. MR Zbl

[18] A. Borovik and Ş. Yalçınkaya, "Black box algebra", book in progress.

[19] A. V. Borovik, M. J. DeBonis, and A. Nesin, "CIT groups of finite Morley rank, I", *J. Algebra* **165**:2 (1994), 258–272. MR Zbl

[20] E. Breuillard, "An exposition of Jordan's original proof of his theorem on finite subgroups of $\mathrm{GL}_n(\mathbb{C})$", *Model Theory* **2**:2 (2023), 429–447. MR

[21] E. Bunina, "Regular bi-interpretability of Chevalley groups over local rings", *Eur. J. Math.* **9**:3 (2023), Paper No. 64, 23. MR Zbl

[22] W. Burnside, "On a class of groups of finite order", pp. 269–276 in *Transactions of the Cambridge Philosophical Society*, vol. XVIII, Cambridge Univ. Press, 1900. Zbl

[23] Y. Chen and C. Shramov, "Automorphisms of surfaces over fields of positive characteristic", preprint, 2022. arXiv 2106.15906v2

[24] G. Cherlin, "Groups of small Morley rank", *Ann. Math. Logic* **17**:1-2 (1979), 1–28. MR Zbl

[25] G. Cherlin, "Good tori in groups of finite Morley rank", *J. Group Theory* **8**:5 (2005), 613–621. MR Zbl

[26] G. Cherlin, "Around the algebraicity problem in odd type", *Model Theory* **3**:2 (2024), 505–538.

[27] G. L. Cherlin and J. Reineke, "Categoricity and stability of commutative rings", *Ann. Math. Logic* **9**:4 (1976), 367–399. MR Zbl

[28] M. J. Collins, "On Jordan's theorem for complex linear groups", *J. Group Theory* **10**:4 (2007), 411–423. MR Zbl

[29] L. J. Corredor and A. Deloro, "A self-contained guide to Frécon's theorem", *Rev. Colombiana Mat.* **55**:2 (2021), 205–210. MR Zbl

[30] L.-J. Corredor, A. Deloro, and J. Wiscons, "Minimal dimensions of $\mathrm{Sym}(n)$- and $\mathrm{Alt}(n)$-modules", in preparation.

[31] C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Pure and Applied Math. **XI**, Interscience Publishers, New York, 1962. MR Zbl

[32] A. Deloro, "Groups of finite Morley rank and their representations", lecture notes, Universidad de los Andes, Bogotá, 2019, available at https://webusers.imj-prg.fr/∼adrien.deloro/teaching-archive/Bogota-Representations.pdf.

[33] A. Deloro, "Zilber's skew-field lemma", *Model Theory* **3**:2 (2024), 571–586.

[34] A. Deloro and E. Jaligot, "Involutive automorphisms of $N_\circ^\circ$-groups of finite Morley rank", *Pacific J. Math.* **285**:1 (2016), 111–184. MR Zbl

[35] A. Deloro and J. Wiscons, "The geometry of involutions in ranked groups with a TI-subgroup", *Bull. Lond. Math. Soc.* **52**:3 (2020), 411–428. MR Zbl

[36] L. E. Dickson, "Representations of the general symmetric group as linear groups in finite and infinite fields", *Trans. Amer. Math. Soc.* **9**:2 (1908), 121–148. MR Zbl

[37] Y. A. Drozd and V. V. Kirichenko, *Finite-dimensional algebras*, Springer, 1994. MR Zbl

[38] K. Erdmann and T. Holm, *Algebras and representation theory*, Springer, 2018. MR Zbl

[39] W. Feit, "Richard D. Brauer", *Bull. Amer. Math. Soc.* (*N.S.*) **1**:1 (1979), 1–20. MR Zbl

[40] O. Frécon, "Simple groups of Morley rank 3 are algebraic", *J. Amer. Math. Soc.* **31**:3 (2018), 643–659. MR Zbl

[41] D. Gorenstein, R. Lyons, and R. Solomon, *The classification of the finite simple groups*, Math. Surveys and Monographs **40**, Amer. Math. Soc., Providence, RI, 1994. MR Zbl

[42] A. Guld, "Finite subgroups of the birational automorphism group are 'almost' nilpotent of class at most two", preprint, 2020. arXiv 2004.11715v1

[43] W. Hodges, "Meeting Boris Zilber", *Model Theory* **3**:2 (2024), 203–211.

[44] M. C. Jordan, "Mémoire sur les équations différentielles linéaires à intégrale algébrique", *J. Reine Angew. Math.* **84** (1878), 89–215. MR Zbl

[45] F. Knop, "Mehrfach transitive Operationen algebraischer Gruppen", *Arch. Math.* (*Basel*) **41**:5 (1983), 438–446. MR Zbl

[46] V. Landazuri and G. M. Seitz, "On the minimal degrees of projective representations of the finite Chevalley groups", *J. Algebra* **32** (1974), 418–443. MR Zbl

[47] M. J. Larsen and R. Pink, "Finite subgroups of algebraic groups", *J. Amer. Math. Soc.* **24**:4 (2011), 1105–1158. MR Zbl

[48] J. G. Loveys and F. O. Wagner, "Le Canada semi-dry", *Proc. Amer. Math. Soc.* **118**:1 (1993), 217–221. MR Zbl

[49] D. Macpherson and A. Pillay, "Primitive permutation groups of finite Morley rank", *Proc. London Math. Soc.* (3) **70**:3 (1995), 481–504. MR Zbl

[50] A. I. Maltsev, "The elementary properties of linear groups", pp. 110–132 in *Certain problems in mathematics and mechanics* (*in honor of M. A. Lavrentev*), Izdat. Sibirsk. Otdel. Akad. Nauk SSSR, Novosibirsk, 1961. In Russian. MR Zbl

[51] Y. Mustafin and B. Poizat, "Sous-groupes superstables de $\mathrm{SL}_2(K)$ et de $\mathrm{PSL}_2(K)$", *J. Algebra* **297**:1 (2006), 155–167. MR Zbl

[52] P. M. Neumann, "The concept of primitivity in group theory and the second memoir of Galois", *Arch. Hist. Exact Sci.* **60**:4 (2006), 379–429. MR Zbl

[53] L. Newelski, "On type definable subgroups of a stable group", *Notre Dame J. Formal Logic* **32**:2 (1991), 173–187. MR Zbl

[54] B. Poizat, "MM. Borel, Tits, Zilber et le général nonsense", *J. Symbolic Logic* **53**:1 (1988), 124–131. MR Zbl

[55] B. Poizat, "Quelques modestes remarques à propos d'une conséquence inattendue d'un résultat surprenant de Monsieur Frank Olaf Wagner", *J. Symbolic Logic* **66**:4 (2001), 1637–1646. MR Zbl

[56] B. Poizat, "La conjecture d'algébricité, dans une perspective historique, et surtout modèle-théorique", *Model Theory* **3**:2 (2024), 479–504.

[57] Y. G. Prokhorov and C. A. Shramov, "Jordan property for the Cremona group over a finite field", *Tr. Mat. Inst. Steklova* **320** (2023), 298–310. In Russian; translated in *Proc. Steklov Inst. Math.* **320**:1 (2023), 278–289. MR Zbl

[58] R. Steinberg, "Representations of algebraic groups", *Nagoya Math. J.* **22** (1963), 33–56. MR Zbl

[59] R. Steinberg, "Chevalley groups", mimeographed notes, Yale University.

[60] J. Tindzogho Ntsiri, "The structure of an SL_2-module of finite Morley rank", *Math. Log. Q.* **63**:5 (2017), 364–375. MR Zbl

[61] A. Wagner, "The faithful linear representation of least degree of S_n and A_n over a field of characteristic 2", *Math. Z.* **151**:2 (1976), 127–137. MR Zbl

[62] A. Wagner, "The faithful linear representations of least degree of S_n and A_n over a field of odd characteristic", *Math. Z.* **154**:2 (1977), 103–114. MR Zbl

[63] F. Wagner, "Subgroups of stable groups", *J. Symbolic Logic* **55**:1 (1990), 151–156. MR Zbl

[64] F. Wagner, "Fields of finite Morley rank", *J. Symbolic Logic* **66**:2 (2001), 703–706. MR Zbl

[65] Ş. Yalçınkaya, "Black box groups", *Turkish J. Math.* **31** (2007), 171–210. MR Zbl

[66] B. I. Zilber, "Groups and rings whose theory is categorical", *Fund. Math.* **95**:3 (1977), 173–188. MR Zbl

[67] B. I. Zilber, "The structure of models of categorical theories and the finite-axiomatizability problem", unpublished manuscript, 1977.

Received 12 Dec 2022. Revised 25 May 2023.

ALEXANDRE BOROVIK:

alexandre@borovik.net
Department of Mathematics, University of Manchester, Manchester, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.571

Zilber's skew-field lemma

Adrien Deloro

We revisit one of Zilber's early results in model-theoretic algebra, viz., definability in Schur's lemma. This takes place in a broader context than the original version from the seventies.

La droite laisse couler du sable.
Toutes les transformations sont possibles.
Paul Éluard

The present contribution discusses and proves a linearisation result originating in Zilber's early work. Let us note to begin:

(1) o-minimal dimension and Borovik–Morley–Poizat rank are examples of finite dimensions.

(2) All necessary definitions are in Section 2.1.

(3) I have preferred not to conflate T with $\mathbb{K}$ in the statement.

(4) There are classical corollaries in Section 2.4.

(5) The result bears no relationship to indecomposable generation discussed in Section 2.5.

Theorem (Zilber's skew-field lemma). *Work in a finite-dimensional theory. Let V be a definable, connected, abelian group and $S, T \leq \mathrm{DefEnd}(V)$ be two invariant rings of definable endomorphisms such that*

- *V is irreducible as an S-module*;
- *$C(S) = T$ and $C(T) = S$, with centralisers taken in* $\mathrm{DefEnd}(V)$;
- *S and T are infinite*;
- *S* ***or*** *T is unbounded.*

Then there is a definable skew-field $\mathbb{K}$ such that $V \in \mathbb{K}\text{-}\mathbf{Vect}_{<\aleph_0}$; moreover, $S \simeq$ $\mathrm{End}(V : \mathbb{K}\text{-}\mathbf{Vect})$ *and $T \simeq \mathbb{K}\,\mathrm{Id}_V$ are definable.*

The present exposition contains results stemming from more general research pursued with Frank O. Wagner [Deloro and Wagner $\geq$ 2024].
MSC2020: primary 03C60, 16B70; secondary 16G99.
Keywords: Schur's lemma, Zilber's field theorem, definability, linearisation.

Section 1 provides context. Section 2 discusses the statement, and gives all definitions. The proof is in Section 3.

1. Introduction

Section 1.1 explains the relation to Schur's Lemma. Section 1.2 makes some historical remarks. Section 1.3 discusses a more famous corollary on fields in abstract groups.

1.1. ***Schur's lemma.*** Among the early work of Zilber are a couple of gems in model-theoretic algebra. (More on Zilber's early work is in [Hodges 2024] in the present volume.) This article deals with one of the phenomena he discovered: *many $\aleph_1$-categorical groups interpret infinite fields*. The result, or the method, or the general line of thought, is often called *Zilber's field theorem*. It stems from Schur's lemma in representation theory:

Lemma (Schur's lemma). *Let R be a ring and V be a simple R-module. Then the covariance ring $\mathbb{F} = C_{\mathrm{End}(V)}(R)$ is a skew-field, V is a vector space over $\mathbb{F}$, and $R \hookrightarrow \mathrm{End}_R(V)$.*

Zilber's deep observation is simple:

in many model-theoretically relevant cases, $\mathbb{F}$ is definable.

A precise and modern form of the latter statement, given as Corollary 1 in Section 2.4, is a straightforward consequence of the main theorem above. (One should remember that every module is actually a bimodule by introducing Schur's covariance ring.) I shall henceforth call it (in long form) the *Schur–Zilber skew-field lemma*, hoping that Boris will not mind being in good company. Far be it from me to minimise its significance by dubbing it a lemma instead of a theorem; quite the opposite as lemmas are versatile devices — methods.

1.2. ***Editorial fortune of the lemma.*** This subsection is a layman's attempt at providing historical remarks. I apologise for misconceptions.

- As one learns from [Curtis 1999, p. 139], Schur's lemma itself appears in [Schur 1904, §2, I.] with comment: "*der auch in der Burnside'schen Darstellung der Theorie eine wichtige Rolle spielt*".

- Before Zilber's result was known, Cherlin [1979, §4.2, Theorem 1] found a definable field independently. There interpretation is obtained by hand (and seemingly by miracle), without a general method. Cherlin heard about Zilber's work after completing his own; [Cherlin 1979, §1.4] is very informative.

• The lemma itself seems not to have drawn as much attention as its corollary on soluble groups (Section 1.3). There are few traces of the lemma as a stand-alone statement.

• All sources discussing the topic [Zilber 1977; 1984; Thomas 1983; Nesin 1989a; 1989b; Poizat 1987; Loveys and Wagner 1993; Borovik and Nesin 1994; Macpherson and Pillay 1995] rely on indecomposable generation (however, see Section 2.5).

• This is different in the o-minimal context, but [Peterzil et al. 2000, Theorem 2.6] has its own techniques. (The earlier [Nesin et al. 1991, Proposition 2.4], which bears no reference to Zilber, resembles the coordinatisation by hand of [Cherlin 1979].) This and the above item may have given the impression that the Schur–Zilber lemma is a finite Morley rank gadget; *the present contribution shows that it isn't*.

• Most sources focus on the ring *generated* by the action instead of going to the centraliser; exceptions are [Nesin 1989a; Macpherson and Pillay 1995]. Only the under-cited [Nesin 1989a] discusses rings and makes the connection with Schur's lemma, while [Macpherson and Pillay 1995, p. 487] notices resemblances between various linearisation results but concludes:

> *There appear to be no immediate implications between this and the results recorded here, though it looks similar to Theorem 1.2.*

The present contribution elucidates the desired relations.

• My own interest in the topic started when I read [Nesin 1989a] while preparing [Deloro 2016]. This resulted in a very partial version of the theorem, in finite Morley rank and using indecomposability. After I gave a talk on generalising "Zilber's field theorem" in Lyon in January 2016, Wagner shared numerous ideas, which will bear all their fruits in the collaboration [Deloro and Wagner $\geq$ 2024].

1.3. ***Fields in soluble groups.*** To some extent, the Schur–Zilber lemma is the poor relation of the following theorem [Zilber 1984, Corollary, p. 175] (currently undergoing generalisation by Wagner):

> *connected, nonnilpotent, soluble groups of finite Morley rank interpret infinite fields.*

I believe the significance of the latter principle has been exaggerated for three reasons.

(1) In the local analysis of simple groups of finite Morley rank, different soluble subquotients may interpret nonisomorphic fields. Since there are strongly minimal structures interpreting *different* infinite fields [Hrushovski 1992], any field structure could be a false lead. (For more on how experts approach the algebraicity conjecture on simple groups of finite Morley rank, and the influence of finite group theory instead of pure model theory, see [Cherlin 2024; Poizat 2024].)

(2) Fields obtained by this method can have "bad" properties, typically nonminimal multiplicative group [Baudisch et al. 2009].

(3) The corollary focused on abstract groups and distracted us from doing representation theory (see the remarkable [Borovik 2024]).

2. The theorem

Section 2.1 contains all necessary definitions. Section 2.2 justifies the structure of the statement. Section 2.3 discusses optimality, Section 2.4 gives corollaries, and Section 2.5 considers the relation to "indecomposable generation".

The general version of the skew-field lemma is a double-centraliser theorem, repeated below. Alternative names could have been "bimodule theorem" or "double-centraliser linearisation".

Theorem. *Work in a finite-dimensional theory. Let V be a definable, connected, abelian group and $S, T \leq \mathrm{DefEnd}(V)$ be two invariant rings of definable endomorphisms such that*

- *V is irreducible as an S-module (viz., in the definable, connected category);*
- *$C(S) = T$ and $C(T) = S$, with centralisers taken in* $\mathrm{DefEnd}(V)$;
- *S and T are infinite;*
- *S **or** T is unbounded.*

Then there is a definable skew-field $\mathbb{K}$ such that $V \in \mathbb{K}\text{-}\mathbf{Vect}_{<\aleph_0}$; moreover, $S \simeq$ $\mathrm{End}(V : \mathbb{K}\text{-}\mathbf{Vect})$ *and $T \simeq \mathbb{K}\,\mathrm{Id}_V$ are definable.*

It would be interesting to recast this kind of double-centraliser result in the abstract ring $S \otimes T$, with no reference to V. (This is not planned in [Deloro and Wagner $\geq$ 2024].)

2.1. *Definitions.*

• Connected: with no definable proper subgroup of finite index. (Since the context does not provide a DCC, not all definable groups have a connected component.)

• Bounded: which does not grow larger when taking larger models. (The algebraist may fix a saturated model with inaccessible cardinality and argue there; bounded then means small. Also see [Halevi and Kaplan 2023].)

• Type-definable: a bounded intersection of definable sets.

• Invariant: a bounded union of type-definable sets. (The name comes from the action of the Galois group of a "large" model. Section 2.2 gives reasons for considering the invariant category instead of the definable one.)

• Irreducible: no nontrivial proper submodule — a submodule being definable *and connected*. (This is weaker than usual algebraic simplicity, which would also exclude finite submodules. Model theory will handle those in its own way.)

• Finite-dimensional: which bears a reasonable dimension on interpretable sets. Here [Wagner 2020] would say *fine*, *integer-valued*, *finite-dimensional*. The definition is as follows.

Definition [Wagner 2020]. A theory T is [fine, integer-valued] finite-dimensional if there is a dimension function dim from the collection of all interpretable sets in models of T to $\mathbb{N} \cup \{-\infty\}$, satisfying the following for a formula $\varphi(x, y)$ and interpretable sets X and Y:

- Invariance: If $a \equiv a'$ then $\dim(\varphi(x, a)) = \dim(\varphi(x, a'))$.
- Algebraicity: X is finite nonempty if and only if $\dim(X) = 0$, and $\dim(\varnothing) = -\infty$.
- Union: $\dim(X \cup Y) = \max\{\dim(X), \dim(Y)\}$.
- Fibration: If $f : X \to Y$ is an interpretable map such that $\dim(f^{-1}(y)) \geq d$ for all $y \in Y$, then $\dim(X) \geq \dim(Y) + d$.

The dimension extends to type-definable, and then to invariant sets; of course one should no longer expect nice additivity properties.

Except for a key "field definability lemma" (Section 2.5) we shall use little from [Wagner 2020]. There is an ACC and a DCC on definable, *connected* subgroups.

2.2. *Explaining the statement.* Our statement deviates from traditional versions in several respects, and we make three cases for three notions.

Skew-fields rather than fields. Schur's lemma produces a skew-field, and so does Zilber's model-theoretic version.

• This went first unnoticed since $\aleph_1$-categorical skew-fields are commutative (answering a question of Macintyre's, proved by Cherlin and Shelah — see note on [Borovik and Nesin 1994, p. 139] — and independently by Zilber [1977].)

• It is easy to construct, in tame geometry, so-called "quaternionic representations", where the Schur field is the skew-field of quaternions.

• Also, the subring $\langle A\rangle \leq \mathrm{End}(V)$ generated by a *commutative* group action can be smaller than its Schur skew-field $C_{\mathrm{End}(V)}(A)$: classical focus on the former (as in most sources) captures only partial geometric information.

So skew-fields are naturally unavoidable. (There remains the question of which skew-fields can arise in a finite-dimensional theory. Skew-fields abound in number theory, but arguably number theory is far from tame. One can also doubt that the more exotic objects constructed in [Cohn 1995] will be finite-dimensional. The bold would conjecture that infinite skew-fields in finite-dimensional theories are commutative and real closed, commutative and algebraically closed, or quaternionic over a commutative real closed field. The more reasonable may be content with conjecturing that they are finite extensions of their centres. Either of these claims, if true, would have an impact on their stability-theoretic properties.)

Rings rather than groups. Let V be an abelian group; then $\mathrm{End}(V)$ is a ring. This accounts for studying representations of *rings*.

• If $G \leq \mathrm{Aut}(V)$ is a definable acting group, the subring of $\mathrm{End}(V)$ it generates need not be definable (see "invariance" below). This may have baffled pioneers in the topic.

• Rings were long neglected after the seminal [Zilber 1977] (a remarkable exception being [Nesin 1989a]). Going to the enveloping ring, however, gives powerful results, inaccessible to group-theoretic reasoning; see [Borovik 2024].

Invariance rather than definability. Leaving definability may have stopped first investigators of the matter; it is however salutary.

• If $G \leq \mathrm{Aut}(V)$ is a definable group, then the generated subring $\langle G\rangle \leq \mathrm{End}(V)$ is $\bigvee$-definable; this is closer to definability than invariance is. However (see "skew-fields" above), $\langle G\rangle$ does not capture enough geometric information. The double-centraliser $C(C(G)) \geq \langle G\rangle$ is more adapted to Schur-style arguments.

• So let $R \leq \mathrm{End}(V)$ be a definable ring. Then Schur's covariance ring $C_{\mathrm{DefEnd}(V)}(R)$ need not be definable, but it is invariant. And if R itself is invariant, $C_{\mathrm{DefEnd}(V)}(R)$ is too.

So model-theoretic invariance arises as naturally as centralisers do.

2.3. ***Optimality.***

• Both S and T must be infinite.

Let $\mathbb{K}$ be a pure algebraically closed field of positive characteristic p and $V = \mathbb{K}_+$, which is definably minimal. Now $\mathrm{DefEnd}(V)$ consists of quasi-p-polynomials, viz., of all maps $x \mapsto \sum_{k=-n}^{n} a_{p^k}\,\mathrm{Fr}_{p^k}$, where Fr is the Frobenius automorphism of relevant power, and $a_{p^k} \in \mathbb{K}$; there is no bound on n. Only the action of $\mathbb{F}_p$ commutes to all these. We then let $S = \mathrm{DefEnd}(V)$ and $T = \mathbb{F}_p$ (or vice-versa). The first is not definable.

• At least one must be unbounded.

For the same V, now let S be the ring of all quasi-p-polynomials *with coefficients in* $\mathbb{F}_p$, viz., the subring of $\mathrm{DefEnd}(V)$ generated by Fr_p and its inverse. Then one easily sees that $C(S) = S$ is countable, and not definable.

On the other hand, it so happens that S-irreducibility can be relaxed to irreducibility as an (S, T)-bimodule [Deloro and Wagner $\geq$ 2024]. So in retrospect, the main theorem can be retrieved as a corollary to [Deloro and Wagner $\geq$ 2024, Theorem 2].

2.4. ***Corollaries.*** I give three corollaries, proved in Section 3.5. The first relates the main, "double-centraliser" theorem to Schur's lemma. The second retrieves what is called "Zilber's field theorem" in sources such as [Borovik and Nesin 1994]. The third is a variation coming from Nesin's work and isolated by Poizat.

Corollary 1 (Schur–Zilber, one-sided form). *Work in a finite-dimensional theory. Let V be a definable, connected, abelian group and $S \leq \mathrm{DefEnd}(V)$ be an invariant, unbounded ring of definable endomorphisms. Suppose that V is irreducible as an S-module. Then $C_{\mathrm{DefEnd}(V)}(S)$ is a definable skew-field.*

Corollary 1 is, however, not equivalent to our main result, which also covers the case of unbounded T and infinite S.

Corollary 2 (see [Deloro 2016, Théorème IV.1]). *Work in a finite-dimensional theory. Let V be a definable, connected, abelian group and $G \leq \mathrm{DefAut}(V)$ be a definable group such that V is irreducible as a G-module* ***and*** *$C_{\mathrm{DefEnd}(V)}(G)$ is infinite. Then $T = C_{\mathrm{DefEnd}(V)}(G)$ is a definable skew-field (so the action of G is linear).*

Corollary 2 (or a minor variation) unifies and should replace various results such as [Zilber 1984, Lemma 2; Loveys and Wagner 1993, Theorem 4; Nesin 1989a, Lemma 12; Macpherson and Pillay 1995, Theorem 1.2(b); Deloro 2016, Théorème IV.1; Peterzil et al. 2000, Theorem 2.6; Macpherson et al. 2000, Proposition 4.1]. However, there are no claims on finite generation.

Corollary 3 (after Nesin and Poizat). *Work in a finite-dimensional theory. Let V be a definable, connected, abelian group and $R \leq \mathrm{DefEnd}(V)$ be an invariant, unbounded, commutative ring of definable endomorphisms. Suppose there is an invariant group $G \leq \mathrm{DefAut}(V)$ such that*

- *V is irreducible as a G-module*;
- *G normalises R*;
- *G is connected.*

Then there is a definable skew-field $\mathbb{K}$ such that $V \in \mathbb{K}\text{-}\mathbf{Vect}_{<\aleph_0}$; moreover, $R \hookrightarrow \mathbb{K}\,\mathrm{Id}_V$ and $G \hookrightarrow \mathrm{GL}(V : \mathbb{K}\text{-}\mathbf{Vect})$.

It would be interesting to relax the assumption on commutativity of R. Further generalisations are expected using endogenies instead of endomorphisms [Deloro and Wagner $\geq$ 2024].

2.5. ***Indecomposable generation (and how to avoid it).*** Contrary to widespread belief, the Schur–Zilber lemma has nothing to do with another celebrated result from Boris' early work: the "indecomposability theorem" [Zilber 1977, Theorem 3.3], which by analogy with the algebraic case I prefer to call the *Chevalley–Zilber* *generation lemma* (again with hope that Boris will not mind being in good company). For more on the topic, see [Poizat 2024, §8].

Both results are often presented jointly, which serves neither clarity nor purity of methods. In contrast, the proof given here relies on another phenomenon.

Lemma (field definability; extracted from [Wagner 2020, Proposition 3.6]). *Work in a finite-dimensional theory. Let $\mathbb{K}$ be an invariant skew-field such that*

- *there is an upper bound on dimensions of type-definable subsets of* $\mathbb{K}$;
- $\mathbb{K}$ *contains an invariant*, *unbounded subset.*

Then $\mathbb{K}$ is definable.

The first clause is satisfied if there is a definable $\mathbb{K}$-vector space of finite $\mathbb{K}$-linear dimension.

3. The proofs

The corollaries are derived in Section 3.5. Let V, S, T be as in the theorem. The proof is a series of claims arranged in propositions.

Proof of Zilber's skew-field lemma. It is convenient to let T act from the right and treat V as an (S, T)-bimodule.

Proposition. (i) *T is a domain acting by surjections with finite kernels*; *for $t \in T \setminus \{0\}$ one has $Vt = V$.*

This will later be reinforced in (x).

Proof. (i) Let $t \in T \setminus \{0\}$. Then $0 < Vt$ is S-invariant, definable, and connected; by S-irreducibility $Vt = V$, so t is onto. In particular, T is a domain. Finally, $\dim \ker t = \dim V - \dim Vt = 0$, so $\ker t$ is finite. □

The global behaviour is difficult to control, so we go down to a more "local" scale with a suitable notion of lines.

3.1. *Lines.*

Notation. Let $\delta = \min\{\dim sV : s \in S \setminus \{0\}\}$ and $\Lambda = \{sV : \dim sV = \delta\}$ be the set of *lines*.

Proposition. (ii) *Every line is T-invariant.*

(iii) *If $L \in \Lambda$ and $s \in S$ are such that $sL \neq 0$, then $sL \in \Lambda$*; *in particular*, *$L \cap \ker s$ is finite.*

(iv) *V is a finite sum of lines.*

(v) *S is transitive on Λ.*

Items (iii) and (iv) will later be reinforced in (vi) and (ix), respectively.

Proof. (ii) This is obvious since S and T commute.

(iii) Say $L = s_0 V$. If $sL \neq 0$, then $0 < \dim sL = \dim((ss_0)V) \leq \dim(s_0 V) = \delta$, so by minimality of δ one has $sL \in \Lambda$. This also implies $\dim(L \cap \ker s) = \dim \ker s_{|L} = \dim L - \dim sL = 0$, and $L \cap \ker s$ is finite.

(iv) The subgroup $0 < \sum \Lambda \leq V$ is definable, connected, and S-invariant; by S-irreducibility, it equals V. Since dimension is finite, it is a finite sum.

(v) Let $L_1, L_2 \in \Lambda$, say $L_i = s_i V$. Now as above, $V = \sum_S sL_1 \not\leq \ker s_2$, so there is $s \in S$ such that $s_2 s L_1 \neq 0$. But then $0 < s_2 s L_1 = s_2 s s_1 V \leq s_2 V = L_2$, and equality holds. □

3.2. *Linearising lines.*

Proposition. (vi) *If $L \in \Lambda$ and $s \in S$ are such that $sL \neq 0$, then $L \cap \ker s = 0$.*

(vii) *T acts by automorphisms on every line.*

The proof is different depending on whether S or T is unbounded.

Proof if T is unbounded. (vi) Suppose $sL \neq 0$; we show $L \cap \ker s = 0$. By (v), S is transitive on Λ, so there is $s' \in S$ with $s'sL = L$. Now $L \cap \ker s \leq L \cap \ker(s's)$, so we may assume that $sL = L$. Recall that $\ker s_{|L} = L \cap \ker s$ is finite by (iii). Considering $s_{|L}^2 : L \to L$, which is onto, we inductively find $|\ker s_{|L}^n| = |\ker s_{|L}|^n$, so $K = \sum_{n \in \mathbb{N}} \ker s_{|L}^n$ is either trivial or countably infinite. *Since T is unbounded*, there is $t \in T \setminus \{0\}$ annihilating K. But t has a finite kernel by (i), so $K = 0$, as desired.

(vii) Let $t \in T$. Then $\ker t$ is finite and S-invariant, while S is infinite; so there is $s_0 \in S \setminus \{0\}$ with $s_0(\ker t) = 0$.

Since $s_0 \neq 0$ and $V = \sum \Lambda$ by (iv), there is $L_0 \in \Lambda$ such that $s_0 L_0 \neq 0$. Then $s_0(L_0 \cap \ker t) = 0$ so $L_0 \cap \ker t \leq L_0 \cap \ker s_0$ by (vi).

Now if L is any other line, then there is $s \in S$ with $sL = L_0$ by (v). Therefore $s(L \cap \ker t) \leq L_0 \cap \ker t = 0$, and $L \cap \ker t \leq L \cap \ker s = 0$ by (vi) again.

So $\ker t$ intersects each line trivially. □

Proof if S is unbounded. The strategy is different here and we first prove weakened versions in reverse order.

Weak (vii)$'$: We first prove that T acts by automorphisms on *some* line. By (iv), $V = \sum \Lambda$ is a finite sum, so there are $L_1, \ldots, L_n$ such that $\bigcap_{i=1}^n \operatorname{Ann}_S(L_i) = 0$. In particular $(S, +) \hookrightarrow \prod_i S/\operatorname{Ann}_S(L_i)$ as abelian groups. *Since S is unbounded*, there exists some line L such that the quotient group $\Sigma = S/\operatorname{Ann}_S(L)$ is unbounded. Let $t \in T \setminus \{0\}$. Then $K = \sum_{n \in \mathbb{N}} \ker t_{|L}^n$ is either trivial or countably infinite. Since Σ is unbounded, there is $\sigma \in \Sigma \setminus \{0\}$ annihilating K, i.e., there is $s \in S$ annihilating K but not L. By (iii) this shows $K = 0$, as desired.

Weak (vi)$'$: We next prove: *if T acts by automorphisms on L*, then for $s \in S$ with $sL \neq 0$ one has $L \cap \ker s = 0$. Indeed, $L \cap \ker s$ is finite by (iii). Since T is infinite there is $t \in T \setminus \{0\}$ with $(L \cap \ker s)t = 0$, but t induces an automorphism of L. This proves (vi), but only for lines on which T acts by automorphisms.

(vii) and (vi): By (vii)$'$, let L be a line on which T acts by automorphisms and L' be another line. Then by transitivity (v), there is $s \in S$ with $sL = L'$. Suppose $w \in L' \cap \ker t$. Then there is $v \in L$ with $sv = w$. Now $s(vt) = (sv)t = wt = 0$, so $vt \in L \cap \ker s = 0$. Since T acts by automorphisms on L, (vi)$'$ implies $v = 0$ and $w = 0$, as desired. □

Since it is unclear at this stage whether every element belongs to a line, we cannot immediately conclude that T acts by automorphisms; this requires writing V as a direct sum.

3.3. ***Globalising local geometries.*** Instead of *morphism of T-modules*, we simply say *T-covariant* map. We tend to reserve it for definable maps, even implicitly.

Proposition. (viii) *Lines are complemented as T-modules*, *viz.*, *for $L \in \Lambda$ there is a definable*, *connected*, *T-invariant $H \leq V$ with $V = L \oplus H$.*

(ix) *V is a finite*, ***direct*** *sum of lines.*

(x) *T is a skew-field acting by automorphisms.*

Proof. (viii) Say $L = s_0 V$. Since $V = \sum_S sL$ by (iv) and (v), there is $s \in S$ with $s_0 s L \neq 0$, so $0 < s_0 s L = s_0 s s_0 V \leq L$. Let $s_1 = s_0 s$, so that $L = s_1 V = s_1 L$. Then for $v \in V$ there is $\ell \in L$ with $s_1 v = s_1 \ell$; in particular, $v = \ell + (v - \ell)$ with $\ell \in L$ and $v - \ell \in \ker s_1$. Therefore $H = \ker s_1$ is such that $V = L + H$; it also is T-invariant as S and T commute. Now $L \cap H = L \cap \ker s_1 = 0$ by (vi), so actually $V = L \oplus H$. Connectedness of H follows.

Since $V = L \oplus H$ is a direct decomposition as a T-module, the associated projections are T-covariant (viz., morphisms of T-modules).

(ix) As long as possible, we recursively construct lines $L_1, \ldots, L_i$ with direct complements H_j (as definable, connected T-modules) satisfying

> for $j \leq i$, one has $L_j \leq \bigcap_{k<j} H_k$ (viz., each new line is contained in all previous complements).

The construction starts by (viii). Now suppose $L_1, \ldots, L_i$ and $H_1, \ldots, H_i$ are as claimed. A quick induction yields:

$$V = \left(\bigoplus_{j=1}^{i} L_j \right) \oplus \left(\bigcap_{j=1}^{i} H_j \right).$$

Let q project V onto $\bigcap_{j=1}^{i} H_i$ with kernel $\bigoplus_{j=1}^{i} L_j$. Then q is T-covariant, so $q \in C(T) = S$. If $\bigoplus_{j=1}^{i} L_j < V$, then $q \neq 0$. Now $V = \sum \Lambda$ so there is $L' \in \Lambda$

such that $qL' \neq 0$. Then let $L_{i+1} = qL' \in \Lambda$; it satisfies $L_{i+1} \leq \bigcap_{j=1}^{i} H_j$. Picking a complement as in (viii), we have reached stage $i+1$.

However the process must terminate because $\dim \bigoplus_{j=1}^{i} L_j = \delta \cdot i$ remains bounded by $\dim V$. So at some stage one obtains $\bigoplus_{j=1}^{i} L_j = V$, as wanted.

(x) Say $V = \bigoplus_{i=1}^{n} L_i$ by (ix). Then for $t \in T$ one has $\ker t = \bigoplus_{i=1}^{n}(L_i \cap \ker t) = 0$ by (vii). $\square$

Hence T is a skew-field and $V \in T$-**Vect**, but we still fall short of definability.

3.4. ***Definability.*** We return to lines. The next result is of a purely auxiliary nature.

Proposition. (xi) *Let $L_1, L_2 \in \Lambda$. If $\sigma : L_1 \simeq L_2$ is definable and T-covariant, then there is an* ***invertible*** *$s \in S^\times$ inducing σ.*

Proof. (xi) Using (viii), write $V = L_1 \oplus H_1$ for some $\pi_1 \in S$ with $L_1 = \operatorname{im} \pi_1$ and $H_1 = \ker \pi_1$.

If $L_2 \cap H_1 = 0$, then H_1 is a common direct complement for L_1 and L_2. Glue $\sigma : L_1 \to L_2$ with Id_H to produce a T-covariant map, viz., an element of $C_{\mathrm{DefEnd}(V)}(T) = S$, inducing σ. It clearly is invertible.

If $L_2 \leq H_1$, then the process proving (ix) enables us to take L_1 and L_2 as the first two lines in a direct sum decomposition. Consider the map given on L_1 by σ, on L_2 by σ^{-1}, and on the remaining sum by 1. It is T-covariant and bijective, hence invertible in S; it induces σ.

The case $0 < L_2 \cap H_1 < L_2$ cannot happen, for then $\ker \pi_{1|L_2} \geq L_2 \cap H_1 > 0$ so by definition of lines, $\pi_1 L_2 = 0$ and $L_2 \leq H_1$. $\square$

Notation. For $L \in \Lambda$, by (viii) there exists a definable, connected, T-invariant H such that $V = L \oplus H$.

- Let π_L be the relevant projection and $S_L = \pi_L S \pi_L$.
- Also let $T_L \leq \mathrm{DefEnd}(L)$ be the image of T.

In full rigour, S_L also depends on the complement chosen; we omit it from the notation. This will not create difficulties.

Proposition. (xii) *S_L and T_L are skew-fields contained in* $\mathrm{DefEnd}(L)$.

(xiii) *Inside* $\mathrm{DefEnd}(L)$ *one has $C(S_L) = T_L$ and $C(T_L) = S_L$.*

(xiv) *T is definable.*

Proof. In case T is unbounded, one may directly jump to (xiv).

(xii) Keep in mind that S_L is an additive subgroup of S closed under multiplication but it need not contain 1. (Sometimes S_L is called a *subrng*, for "subring without identity".) However, S_L per se is a ring with identity π_L, as the latter acts on L as Id_L. Moreover, if $\pi_L s \pi_L$ annihilates L, then since it annihilates the chosen direct complement, it is 0 as an endomorphism of V, viz., $\pi_L s \pi_L = 0$ in S. So S_L can be viewed as a subring of $\mathrm{DefEnd}(L)$, and it is exactly the subring of restrictions-corestrictions $\{s_{|L}^{|L} : s \in \mathrm{Stab}_S(L)\}$. (This explains why the complement plays no role in our construction. It is however useful to have both points of view on S_L.)

Let $s \in S_L \setminus \{0\}$. Then $sL = L$, so by (vi) and since S and T commute, it induces some T-covariant automorphism σ of L; by (xi) there is $s' \in S^\times$ inducing σ. Now $\pi_L s'^{-1} \pi_L$ is a two-sided inverse of s in S_L. This proves that S_L is a skew-field. So is T by (x); now the restriction map $T \to T_L$, which is onto by definition, is injective since T acts by automorphisms. Therefore T_L is a skew-field as well.

(xiii) One of them is easy. Let $f : L \to L$ be a definable, T_L-covariant morphism, viz., $f \in C_{\mathrm{DefEnd}(L)}(T_L)$. By definition, f commutes with the action of T. Take any T-invariant direct complement H and set $\hat{f} = 0$ on H. Then $\hat{f} : V \to V$ is T-covariant. Hence $\hat{f} \in C(T) = S$ and $\pi_L \hat{f} \pi_L = f \in S_L$.

Now let $g : L \to L$ be definable and S_L-covariant, viz., $g \in C_{\mathrm{DefEnd}(L)}(S_L)$. We aim at extending g to an S-covariant endomorphism of V.

For $M \in \Lambda$ first use transitivity (v) to choose $s \in S$ with $sL = M$. By (xi) we may assume $s \in S^\times$. Notice that sgs^{-1} leaves M invariant, and let $g_M \in \mathrm{DefEnd}(M)$ be the induced map. We claim that this does not depend on the choice of s. Indeed let s' be another invertible choice, giving rise to g'_M. Then $s^{-1}s'$ induces an element of S_L, so g commutes with it and we find $g_M = g'_M$.

We deduce as follows that $g_M \in C(S_M)$. For if $\eta \in S_M$ then we may assume $\eta \neq 0$ so by (xi) it is induced by an invertible element $h \in S^\times$ normalising M. Then $s' = hs$ is another invertible element taking L to M. By the preceding paragraph, $s'gs'^{-1} = hg_M h^{-1}$ and $sgs^{-1} = g_M$ agree on M, so g_M commutes with η in the ring S_M.

We even prove: if $s \in S$ induces $\sigma : M \simeq N$, then $g_N \sigma = \sigma g_M$. Both are maps from M to N. By (xi), we freely suppose s invertible and pick invertible s_M, s_N inducing $L \simeq M, N$. Then $s'_M = s^{-1}s_N \in S$ takes L to M, so $s'_M g s'^{-1}_M$ agrees with $s_M g s_M^{-1} = g_M$ on M. Thus for arbitrary $m \in M$ we find

$$\begin{aligned} g_N \sigma(m) &= ss^{-1} \cdot s_N g s_N^{-1} \cdot s(m) \\ &= s \cdot (s^{-1}s_N) g (s_N^{-1}s)(m) = s g_M(m) = \sigma g_M(m). \end{aligned}$$

Therefore $g_N \sigma = \sigma g_M$, as claimed.

Finally take a direct sum $V = \bigoplus L_i$ as in (ix) and let $\hat{g}\left(\sum \ell_i\right) = \sum g_{L_i}(\ell_i)$, which is definable, well-defined, and extends g. We want to show $\hat{g} \in C(S)$. Let $s \in S$; also let $s_i = \pi_i s$. It is enough to show that $\hat{g}$ commutes with each s_i, and it is enough to show that they commute on each L_j. We have thus reduced to checking that $\hat{g}$ and $\sigma : L_j \simeq L_i$ induced by an element of S commute. But this is the previous paragraph.

Hence $\hat{g} \in C(S) = T$ and therefore $g = \hat{g}_{|L} \in T_L$.

(xiv) Recall that T is a skew-field by (x). If T is unbounded we directly apply the field definability lemma from Section 2.5 (in that case, (xii) and (xiii) are not necessary). So we suppose that S is unbounded.

We first prove that there is L such that S_L is unbounded. By (ix) take any decomposition $V = \bigoplus_{i=1}^n L_i$ and form projections π_i onto L_i with kernels $\bigoplus_{j \neq i} L_j$. Let $S_{i,j} = \pi_i S \pi_j$, an additive subgroup of S. We contend that one of them is unbounded. Indeed, the additive group homomorphism

$$S \to \prod_{i,j} S_{i,j}, \quad s \mapsto (\pi_i s \pi_j)_{i,j},$$

is injective since $\sum_k \pi_k = 1$. Now if $S_{L,M}$ and $S_{L',M'}$ are defined as the $S_{i,j}$, one easily sees $S_{L,M} \simeq S_{L',M'}$ definably; so all rings S_L are unbounded.

A caveat: because S_L and T_L are mutual centralisers only in $\mathrm{DefEnd}(L)$ and not in $\mathrm{End}(L)$, the following paragraph cannot be made more trivial.

Therefore S_L is an unbounded skew-field by (xii). By field definability of Section 2.5, S_L is definable; now $\dim S_L > 0$ and $\dim L$ is finite, so $L \in S_L\text{-}\mathbf{Vect}_{<\aleph_0}$. In particular, all S_L-endomorphisms of L are definable, so by (xiii) one has $T_L = \mathrm{End}(L : S_L\text{-}\mathbf{Vect})$. This is a skew-field by (xii), so the *linear* dimension over S_L is 1 and $T \simeq T_L \simeq S_L^{\mathrm{op}}$ is unbounded as well. $\square$

By field definability, the skew-field T is definable and infinite, so $\dim T > 0$; now $\dim V$ is finite so $V \in T\text{-}\mathbf{Vect}_{<\aleph_0}$. Finally $S = C(T) = \mathrm{End}(V : T\text{-}\mathbf{Vect})$. Lines in our sense now coincide with 1-dimensional T-subspaces of V. This completes the proof of Zilber's skew-field lemma. $\square$

3.5. ***Proofs of corollaries.*** We repeat the statements already given in Section 2.4.

Corollary 1 (Schur–Zilber, one-sided form). *Work in a finite-dimensional theory. Let V be a definable, connected, abelian group and $S \leq \mathrm{DefEnd}(V)$ be an invariant, unbounded ring of definable endomorphisms. Suppose that V is irreducible as an S-module. Then $C_{\mathrm{DefEnd}(V)}(S)$ is a definable skew-field.*

Proof. Let $T = C_{\mathrm{DefEnd}(V)}(S)$. Notice that T acts by surjective endomorphisms, so it is a domain. If it is finite, then it is a definable field. Otherwise we wish to apply our theorem, but it is unclear whether $S = C_{\mathrm{DefEnd}(V)}(T)$. It actually does not matter. Let $\hat{S} = C_{\mathrm{DefEnd}(V)}(T) \geq S$, which is invariant and unbounded. Moreover, $C_{\mathrm{DefEnd}(V)}(\hat{S}) = T$ as a "triple centraliser", and V remains $\hat{S}$-minimal. So we apply the theorem with $(\hat{S}, T)$ and get definability of the skew-field $C_{\mathrm{DefEnd}(V)}(\hat{S}) = T$. $\square$

Corollary 2. *Work in a finite-dimensional theory. Let V be a definable, connected, abelian group and $G \leq \mathrm{DefAut}(V)$ be a definable group such that V is irreducible* *as a G-module* **and** $C_{\mathrm{DefEnd}(V)}(G)$ *is infinite. Then* $T = C_{\mathrm{DefEnd}(V)}(G)$ *is a definable skew-field* (*so the action of* G *is linear*).

Proof. Let $T = C_{\mathrm{DefEnd}(V)}(G)$ and $S = C_{\mathrm{DefEnd}(V)}(T) \supseteq G$. Apply the theorem. □

Corollary 3 (after Nesin and Poizat). *Work in a finite-dimensional theory. Let* V *be a definable*, *connected*, *abelian group and* $R \leq \mathrm{DefEnd}(V)$ *be an invariant*, *unbounded*, *commutative ring of definable endomorphisms. Suppose there is an invariant group* $G \leq \mathrm{DefAut}(V)$ *such that*

- V *is irreducible as a* G*-module*;
- G *normalises* R;
- G *is connected.*

Then there is a definable skew-field $\mathbb{K}$ *such that* $V \in \mathbb{K}\text{-}\mathbf{Vect}_{<\aleph_0}$; *moreover*, $R \hookrightarrow \mathbb{K}\,\mathrm{Id}_V$ *and* $G \hookrightarrow \mathrm{GL}(V : \mathbb{K}\text{-}\mathbf{Vect})$.

Proof. Let V, R, G be as in the statement. The proof follows that of [Poizat 1987, Théorème 3.8] closely. Let $W \leq V$ be R-irreducible, viz., minimal as a definable, connected, R-submodule; this exists by the DCC on definable, connected subgroups. Let $\mathfrak{p} = \mathrm{Ann}_R(W)$, a relatively definable ideal of R.

For $g \in G$, the definable, connected subgroup $gW \leq V$ is R-invariant, and hence an R-submodule. Clearly $\mathrm{Ann}_R(gW) = g\mathfrak{p}g^{-1}$. Moreover, $R/\mathfrak{p} \simeq R/(g\mathfrak{p}g^{-1})$.

Now, by G-irreducibility, $V = \sum_G gW$. So there are $g_1, \ldots, g_n \in G$ such that $V = \sum_{i=1}^n g_i W$. In particular, $\bigcap_{i=1}^n \mathrm{Ann}_R(g_i W) = 0$, and $R \hookrightarrow \prod R/(g_i \mathfrak{p} g_i^{-1})$. We just saw that all terms have the same cardinality. They are therefore unbounded.

Hence, the unbounded, commutative ring $R/\mathfrak{p}$ acts faithfully on the $R/\mathfrak{p}$-irreducible module W. Notice that $R/\mathfrak{p} \leq C_{\mathrm{DefEnd}(W)}(R/\mathfrak{p})$. By the theorem, the action of $R/\mathfrak{p}$ on W is linearisable, and $R/\mathfrak{p}$ acts by scalars. The problem is to make this linear structure global without losing the action of G. But we know that $\mathfrak{p}$ is a prime ideal of R.

Now consider the set of prime ideals $P = \{h\mathfrak{p}h^{-1} : h \in G\}$. Suppose $\mathfrak{p}_1, \ldots, \mathfrak{p}_k \in P$ are distinct, say $\mathfrak{p}_i = h_i \mathfrak{p} h_i^{-1}$. By prime avoidance, there are elements $r_i \in \mathfrak{p}_i \setminus \bigcup_{j \neq i} \mathfrak{p}_j$. Then taking products, there are elements $r_i' \in \bigcap_{j \neq i} \mathfrak{p}_j \setminus \mathfrak{p}_i$. These are used to show that the sum $\sum_{i=1}^k h_i W$ is direct. In particular, $k \leq \dim V$ and P is finite.

Since G is connected and transitive on the finite set P, the latter is a singleton, namely $P = \{\mathfrak{p}\}$. But by faithfulness one had $\bigcap P = 0$, so $\mathfrak{p} = 0$.

Now let $r \in R \setminus \{0\}$. Then $r \notin \mathfrak{p}$ acts on W as a nonzero scalar, so $W \leq \mathrm{im}\, r$. Since r was arbitrary, for any $g \in G$, one has $gW \leq \mathrm{im}\, r$. Summing, $\mathrm{im}\, r = V$; this implies that $\ker r$ is finite. Then $K = \sum_{n \in \mathbb{N}} \ker r^n$ is either trivial or countably infinite. But by commutativity, it is R-invariant. Since R is unbounded, there is $r_0 \in R \setminus \{0\}$ annihilating K. Since r_0 has a finite kernel in V, we see $K = 0$. Thus the domain R acts by automorphisms on V.

Hence $\mathbb{F} = \mathrm{Frac}(R)$ is naturally a subring of $\mathrm{DefEnd}(V)$. By field definability, it is definable. Now G normalises $\mathbb{F}$ and centralises it [Wagner 2020, §3.3]. In particular, G centralises R. Therefore, $S = C_{\mathrm{DefEnd}(V)}(R)$, which contains R by commutativity, also contains G. It follows that V is S-irreducible and we apply the theorem globally to conclude. $\square$

Acknowledgements

It is a pleasure to thank, alphabetically: Tuna Altınel, who first taught me Zilber's field theorem; Alexandre Borovik and Gregory Cherlin, who took me further into model-theoretic algebra; Ali Nesin and Bruno Poizat, for sharing their culture and passion; Frank Wagner, to whom I have a great technical debt; and last but not least, for a long-lasting vision which inspired so many of us, Boris Zilber.

References

[Baudisch et al. 2009] A. Baudisch, M. Hils, A. Martin-Pizarro, and F. O. Wagner, "Die böse Farbe", *J. Inst. Math. Jussieu* **8**:3 (2009), 415–443. MR Zbl

[Borovik 2024] A. Borovik, "Finite group actions on abelian groups of finite Morley rank", *Model Theory* **3**:2 (2024), 539–569.

[Borovik and Nesin 1994] A. Borovik and A. Nesin, *Groups of finite Morley rank*, Oxford Logic Guides **26**, Oxford Univ. Press, 1994. MR Zbl

[Cherlin 1979] G. Cherlin, "Groups of small Morley rank", *Ann. Math. Logic* **17**:1-2 (1979), 1–28. MR Zbl

[Cherlin 2024] G. Cherlin, "Around the algebraicity problem in odd type", *Model Theory* **3**:2 (2024), 505–538.

[Cohn 1995] P. M. Cohn, *Skew fields*, Encyclopedia of Mathematics and its Applications **57**, Cambridge Univ. Press, 1995. MR Zbl

[Curtis 1999] C. W. Curtis, *Pioneers of representation theory: Frobenius, Burnside, Schur, and Brauer*, History of Mathematics **15**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

[Deloro 2016] A. Deloro, *Un regard élémentaire sur les groupes algébriques*, mémoire d'habilitation, Université Pierre et Marie Curie, 2016, available at https://hal.science/tel-01989901.

[Deloro and Wagner $\geq$ 2024] A. Deloro and F. O. Wagner, "Endogenies and linearisation (working title)", in preparation.

[Halevi and Kaplan 2023] Y. Halevi and I. Kaplan, "Saturated models for the working model theorist", *Bull. Symb. Log.* **29**:2 (2023), 163–169. MR Zbl

[Hodges 2024] W. Hodges, "Meeting Boris Zilber", *Model Theory* **3**:2 (2024), 203–211.

[Hrushovski 1992] E. Hrushovski, "Strongly minimal expansions of algebraically closed fields", *Israel J. Math.* **79**:2-3 (1992), 129–151. MR Zbl

[Loveys and Wagner 1993] J. G. Loveys and F. O. Wagner, "Le Canada semi-dry", *Proc. Amer. Math. Soc.* **118**:1 (1993), 217–221. MR Zbl

[Macpherson and Pillay 1995] D. Macpherson and A. Pillay, "Primitive permutation groups of finite Morley rank", *Proc. London Math. Soc.* (3) **70**:3 (1995), 481–504. MR Zbl

[Macpherson et al. 2000] D. Macpherson, A. Mosley, and K. Tent, "Permutation groups in o-minimal structures", *J. London Math. Soc.* (2) **62**:3 (2000), 650–670. MR Zbl

[Nesin 1989a] A. Nesin, "Nonassociative rings of finite Morley rank", pp. 117–137 in *The model theory of groups* (Notre Dame, IN, 1985–1987), edited by A. Nesin and A. Pillay, Notre Dame Math. Lectures **11**, Univ. Notre Dame Press, 1989. MR Zbl

[Nesin 1989b] A. Nesin, "Solvable groups of finite Morley rank", *J. Algebra* **121**:1 (1989), 26–39. MR Zbl

[Nesin et al. 1991] A. Nesin, A. Pillay, and V. Razenj, "Groups of dimension two and three over o-minimal structures", *Ann. Pure Appl. Logic* **53**:3 (1991), 279–296. MR Zbl

[Peterzil et al. 2000] Y. Peterzil, A. Pillay, and S. Starchenko, "Simple algebraic and semialgebraic groups over real closed fields", *Trans. Amer. Math. Soc.* **352**:10 (2000), 4421–4450. MR Zbl

[Poizat 1987] B. Poizat, *Groupes stables*, Nur al-Mantiq wal-Ma'rifah **2**, Bruno Poizat, Villeurbanne, 1987. MR Zbl

[Poizat 2024] B. Poizat, "La conjecture d'algébricité, dans une perspective historique, et surtout modèle-théorique", *Model Theory* **3**:2 (2024), 479–504.

[Schur 1904] I. Schur, "Neue Begründung der Theorie der Gruppencharaktere", *Berl. Ber.* (1904), 406–432. Zbl

[Thomas 1983] S. R. Thomas, *Classification theory of simple locally finite groups*, Ph.D. thesis, Bedford College, University of London, 1983, available at https://www.proquest.com/docview/1791501281. MR

[Wagner 2020] F. O. Wagner, "Dimensional groups and fields", *J. Symb. Log.* **85**:3 (2020), 918–936. MR Zbl

[Zilber 1977] B. I. Zilber, "Группы и кольца, теория которых категорична", *Fund. Math.* **95**:3 (1977), 173–188. MR Zbl

[Zilber 1984] B. I. Zilber, "Some model theory of simple algebraic groups over algebraically closed fields", *Colloq. Math.* **48**:2 (1984), 173–180. MR Zbl

Received 14 Dec 2022. Revised 15 Oct 2023.

ADRIEN DELORO:

adrien.deloro@imj-prg.fr
Institut de Mathématiques de Jussieu, Sorbonne Université, Paris, France

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.587

Zilber–Pink, smooth parametrization, and some old stories

Yosef Yomdin

The Zilber–Pink conjecture pertains to the "finiteness of unlikely intersections" and falls within the realms of logic, algebraic, and arithmetic geometry. Smooth parametrization involves dividing mathematical objects into simple pieces and then representing each piece parametrically while maintaining control over high-order derivatives. Originally, such parametrizations emerged and were predominantly utilized in applications of real algebraic geometry in smooth dynamics.

The paper comprises two parts. The first part provides informal insights into certain basic results and observations in the field, aimed at elucidating the recent convergence of the seemingly disparate topics mentioned above. The second part offers a retrospective account spanning from 1964 to 1974. During that period, Boris and I studied at the same places, initially in Tashkent and later in Novosibirsk Akademgorodok.

1. Introduction

The author first encountered Boris Zilber at the 110th Tashkent Physics-Mathematics School in 1964. From then until 1974, we shared the same academic journey, studying mathematics in Tashkent, and later in Novosibirsk Akademgorodok. While we engaged in numerous discussions about mathematics, our paths diverged in terms of specialization: Boris delved into mathematical logic and model theory, while I pursued analysis and differential topology. Initially, the gap seemed immense. However, mathematics is a unified discipline! It is one! After many years, my favorite topic, *smooth parametrization*, emerged as an important tool in the recent remarkable progress in the Zilber–Pink conjecture.

In Sections 2 to 4 below, I attempt to explain, in a very informal manner, the connections between these seemingly distinct topics. I am grateful to have received insights from some of the most active participants in the modern research towards the Zilber–Pink conjecture. Their explanations were indispensable to me. I hope that my brief presentation below can be of assistance to some readers.

Finally, in Section 5, I share some recollections from the Tashkent and Akademgorodok period, from 1964 to 1974, which Boris and I experienced together.

MSC2020: 03C64.
Keywords: Zilber–Pink, smooth parametrization, biography.

2. The Zilber–Pink conjecture, and how one can prove it

The Zilber–Pink conjecture ([7; 20; 28], see also [17; 27]) concerns "unlikely intersections". The intersection of two algebraic subvarieties in a variety V is deemed unlikely if its dimension is larger than expected. The conjecture asserts that under some conditions the sets of unlikely intersections are finite. The conjecture offers a uniform framework for various classical problems in Algebraic and Arithmetic Geometry, along with other significant consequences. Recently, there has been dramatic progress in several classical problems, directly related to the Zilber–Pink conjecture: [2; 4; 19; 21] is a very small sample.

Various forms of "smooth parametrization" have played an important role in this progress. Before delving into this topic in Section 4, let's provide a highly informal and intuitive overview of the Pila–Zannier approach that has recently enabled the proof of some very important specific versions of the Zilber–Pink conjecture.

In fact, in the applications of the Pila–Zannier strategy to special point or unlikely intersection problems, the points v which Condition A below concerns, live in a certain preimage $\tilde{V}$ of the algebraic V, rather than in V itself.

In many cases a "height" can be associated to the objects $v \in \tilde{V}$ we aim to count in Zilber–Pink. For a rational number $r = p/q$ the height $H(r)$ is defined as $\max\{|p|, |q|\}$. Similarly for torsion points on pseudo-abelian varieties, and so on. Suppose the number of $v \in \tilde{V}$ with $H(v) \leq H$ is finite for each $H < \infty$, and let $N(\tilde{V}, H)$ denote the cardinality of the set $v \in \tilde{V}$, with $H(v) \leq H$. We then assume the following Condition A: *For a transcendental $\tilde{V}$ and for each $\epsilon > 0$ there exists a constant $c(\epsilon)$ such that*

$$N(\tilde{V}, H) \leq c(\epsilon) H^{\epsilon}, \quad H > 0. \tag{2.1}$$

Results of this nature are now available for counting rational points on transcendental varieties, and in many other cases, starting with the fundamental works of Bombieri and Pila [6] and Pila and Wilkie [18]. Smooth parametrizations appeared, in the context of Diophantine geometry, essentially, in [6]. We discuss them, in somewhat more detail, in Section 4.

Now let us make an additional assumption, called Condition B: *There exist $\epsilon_0 > 0$ and $C > 0$ such that for any $H > 0$ if there are v of height H then, in fact,*

$$N(\tilde{V}, H) \geq C H^{\epsilon_0}. \tag{2.2}$$

In some important cases this second assumption is also satisfied (for instance, due to the Galois group action on the v's; see [4; 21]). Now, if Conditions A and B are satisfied, we get the required finiteness. Indeed, fix $\epsilon < \epsilon_0$ in (2.1). If there exist v's with an arbitrarily big height H, we get a contradiction with the asymptotic bound, for $H \to \infty$, provided by (2.2). We conclude that the height of v is bounded, and hence, so is the total number of v's. This completes the sketch of the Pila–Zannier approach to the Zilber–Pink conjecture.

3. Bombieri–Pila and Pila–Wilkie

In this section we provide a very informal overview of some results and approaches presented in the foundational papers by Bombieri–Pila [6] and Pila–Wilkie [18]. These papers establish, among other things, Condition A for counting rational points on certain transcendental analytic varieties. In [6], the focus is on curves, while [18] extends the results to varieties of higher dimension, definable in a certain o-minimal structure.

The approach involves considering rational points v of a given height H on the graph Γ of a C^k-function Ψ with explicitly bounded derivatives up to order k. The objective is to demonstrate that all these rational points lie on a "small" number of algebraic hypersurfaces W of a certain degree d, which depends on the dimension, k, and H. Later, a form of Bézout's theorem, if available, is utilized to bound the intersections $\Gamma \cap W$, which contain our rational points v.

The key step in this approach is to derive an upper bound on certain Vandermonde-type determinants *VdM*, whose vanishing indicates that the points v lie on an algebraic hypersurface W of degree d. Here the k-smoothness of Ψ and Γ comes into play: the entries of the *VdM* are represented via Taylor expansion, leading to significant cancellations, and ultimately, to the required upper bound.

On the other hand, since the entries of the *VdM* are rational points v of the given height H, the determinants *VdM* are themselves rational numbers with the height, explicitly bounded by a certain D, which depends on H, the number of points, and the dimensions. Therefore, if we can show that $|VdM| < \frac{1}{D}$, we conclude that, indeed, $VdM = 0$, implying that our points lie on an algebraic hypersurface W of degree d. Orchestrating the interrelations between H, k, d, and other parameters is a highly nontrivial task, but ultimately successful.

This concludes the process of counting rational points of a given height H on the graph Γ of a C^k-function Ψ with explicitly bounded derivatives up to order k.

The paper [16] was useful to the author in better understanding (especially in several variables) this part of the approach of [6].

To extend the result from a smooth piece Γ to counting rational points on a transcendental analytic variety V, it becomes necessary to cover V with the graphs Γ_j of C^k-functions Ψ_j, with explicitly bounded derivatives up to order k. Such a covering is what we refer to as a *smooth parametrization*. The existence of such smooth parametrizations for V — a bounded semialgebraic set — was demonstrated in [22] and [14]. While this result sufficed for applications in dynamics, for which it was initially intended, it needed to be extended to analytic varieties for applications in counting rational points.

This extension of smooth parametrizations to sets V definable in a certain o-minimal structure was achieved in [18], together with proving condition A for such V: for each $\epsilon > 0$,

$$N(V^{\mathrm{tr}}, H) \le c(\epsilon) H^{\epsilon}, \quad H > 0,$$

where V^{tr} denotes the "transcendental part" of V. Moreover, the very important *Wilkie conjecture* was posed in [18]: if V is definable in the o-minimal structure, generated by the exponential function, then, in fact, $N(V^{\mathrm{tr}}, H)$ is bounded by a polynomial in $\log H$.

Some special cases of the Wilkie conjecture were settled before a restricted case of this conjecture was proved in [1]. Finally, the full conjecture was confirmed in [5]. New important developments in o-minimal structures and in smooth parametrizations were achieved in [5], a paper that also offers an excellent introduction to smooth parametrizations.

4. Smooth parametrizations

In this section we discuss smooth parametrizations in somewhat more detail. We include a short and informal discussion of the striking recent work [2], where a powerful new class of analytic parametrizations was defined.

"Parametrization" is a change of variables that simplifies the understanding of a mathematical structure under investigation. The most important example in the realm of algebraic and analytic geometry is provided by the resolution of singularities, in its various versions. In many problems of dynamics, analysis, Diophantine and computational geometry it is crucial to maintain control over high-order derivatives while performing a change of variables. Parametrizations of this type are referred to as "smooth parametrizations".

An illustrative example is provided by the C^k*-parametrization* of a semi-algebraic set A. This can be seen as a high-order quantitative version of the well-known result on the existence of a triangulation of such sets A, with the number of simplices bounded in terms of the combinatorial data (the *degree*) of A. In a C^k-version we additionally require that each simplex S_j in the triangulation be an image of the standard simplex Δ, under the parametrization mapping Ψ_j, with all the derivatives of Ψ_j up to order k uniformly bounded.

To state the C^k-parametrization theorem of [14; 23; 22] more precisely, let's recall the definition of semi-algebraic sets. Semi-algebraic sets in $\mathbb{R}^n$ are defined by a finite number of real polynomial equations and inequalities, plus set-theoretic operations.

Given a semi-algebraic set $A \subset \mathbb{R}^n$, the diagram $D(A)$ of the set A comprises the "discrete" data of A — the ambient dimension n, the degrees and the number of the equations and inequalities, and the set-theoretic formula defining A. Hence, $D(A)$ does not depend on specific values of the coefficients of the polynomials involved.

Theorem 4.1. *For any natural k and for any compact semi-algebraic set A inside the cube I^n in $\mathbb{R}^n$, there exists a C^k-parametrization of A, with the number $N(A,k)$ of the C^k-charts Ψ_j, depending only on k and on the diagram $D(A)$ of A.*

The bound on $N(A,k)$ obtained in [14; 23; 22; 8] was explicit but high (initially doubly exponential in k). See also [3]. While this sufficed for the intended applications in dynamics (the "entropy conjecture" for C^∞ maps), it soon became apparent that controlling questions like the semicontinuity modulus of the entropy required polynomial growth of $N(A,k)$ in k. This remained an open problem for a long time, along with some dynamical consequences.

To circumvent these difficulties, *analytic parametrizations* were introduced in [24]. Here, we require the above parametrization mappings Ψ_j to be real analytic, extendible to a complex neighborhood of Δ_j of a controlled size, and explicitly bounded there. This worked in dimensions 1 (and also 2, for diffeomorphisms), but faced challenges in higher dimensions. The primary issue was that typically, an infinite number of analytic charts Ψ_j was required to cover A, because of the hyperbolic geometry of the problem. This direction was further developed in [24; 25; 11; 12; 13], but the finiteness problem remained unsolved.

Let us mention also [15], where some initial steps towards applications of smooth parametrization in computational geometry were provided, and [26], which gave an overview of different types of smooth parametrization and their possible applications (up to 2015).

As for newer advances, let us mention, besides [2], a very recent development by D. Burguet [9] of smooth parametrization techniques for dynamics of curves. It led to the solution of long-standing open problems in smooth dynamics.

Finally, in [2] a new type of analytic parametrization was introduced and termed *complex cellular structures*. The key distinction between complex cellular structures and analytic parametrizations is that the domain of the parametrization mappings Ψ_j (in complex dimension one) is either the unit disk, as before, *or an annulus with a prescribed ratio between radii*. In higher dimensions the domains of the parametrization mappings Ψ_j are constructed inductively, combining the two one-dimensional models. The construction and proofs heavily rely on complex hyperbolic geometry.

Complex cellular structures not only restored the finiteness of parametrizations but achieved much more. In particular, a polynomial bound on the grows of $N(A,k)$ with k in Theorem 4.1 was established, thus proving several longstanding conjectures in smooth dynamics (in combination with [10]). Complex cellular structures provide significant advances also in Diophantine geometry.

Anticipated further progress in these areas is highly promising, and likely to address also open questions regarding various types of smooth and analytic parametrizations, including those raised in [24; 25; 11; 12; 13].

5. The old stories

I first met Boris at the 110th Tashkent Physics-Mathematics School in 1964, if I remember correctly. This school was a remarkable place to learn, to hope, and to dream. Mathematics, in the form of problems to solve, books, lectures, and discussions, was omnipresent. One of our schoolmates once proposed a solution to the Fermat problem ($x^k + y^k = z^k$), and I (like many others, I believe) could not sleep until finding a flaw in the colleague's arguments. Then there were mathematical Olympiads, starting with the school level, then advancing to the city level, and finally reaching the All-Siberian Olympiad in 1965, held at Novosibirsk State University in the famous Akademgorodok near Novosibirsk. Both Boris and I were among the winners of the lower level Olympiads and were invited to the All-Siberian Olympiad.

However, before we arrived (in August 1965), an unusual incident occurred. On the first of May 1965, as on any other May Day, there was a mass demonstration organized by the authorities at the central square of Tashkent, the capital of Uzbekistan, then a part of the USSR. All the glory and power of Uzbekistan's authorities were showcased at the central podium. We, at our 110th school, were compelled to participate in this mass demonstration. As per tradition, when the columns of participants passed near the central podium, the loudspeakers usually announced congratulations and greetings to the Communist Party, the people of the Soviet Union, or other similarly grand entities. Sometimes, however, the congratulations were more specific, such as to the workers of a particular industrial plant currently passing near the podium. And as we passed, a miracle occurred: the loudspeakers congratulated the winners of the preliminary tour of the All-Siberian Olympiad, explicitly mentioning our humble names!

We, in our small group, were elated and proud, but this was not the end of the story. Comrade Rashidov, the first secretary of the Uzbek Communist Party, who was present at the central podium, immediately noticed that among the six explicitly mentioned names of the winners, there was no clear Uzbek name! Perhaps one of our good friends had a partial Uzbek heritage. However, even this winner, upon investigation, turned out to be only half Uzbek and half Tatar. Comrade Rashidov promptly demanded correction of this egregious error. The next day, as usual, the central Tashkent Russian-language newspaper *Pravda Vostoka* (something like “The Truth of the East”) published a detailed report on the May Day 1965 mass demonstration in Tashkent. Included in the newspaper were the names of the winners of the preliminary tour of the All-Siberian Olympiad: six distinctly Uzbek names that I was hearing for the first time. Our small group was a little apprehensive — were we still to go to the All-Siberian Olympiad? But no specific instructions to the contrary followed, so we decided to proceed as if everything were in order.

Comrade Rashidov was, of course, not the first to correct, in line with Party directives, minor personal matters. There was a similar incident involving Stalin. Once, he was quite displeased with certain verbal statements made by Lenin’s widow, Nadezhda Krupskaya. Stalin ordered his subordinates: “Tell this fool that if she does not cease, we will find another widow for Comrade Lenin.” But you see, by 1965, Stalin’s era had firmly passed! It was our original group that eventually made it to the All-Siberian Olympiad.

In total, six of us were invited, all from the 110th Tashkent school. We embarked on the three-day train journey from Tashkent to Novosibirsk, accompanied by our math teacher, Tamara Vladimirovna Reshetnikova.

The three weeks at the Summer School in Akademgorodok, which included the final stage of the All-Siberian Olympiad, were truly exhilarating! Both Boris and I were among the winners of the final stage, granting us an opportunity to enroll in Novosibirsk Physics-Mathematics School. I decided to seize this exciting opportunity, while Boris opted to return to the Tashkent 110th school. However, a year later, he returned to Akademgorodok to participate in the entrance examinations to Novosibirsk State University. I was also there; despite finishing with top grades at the Novosibirsk Physics-Mathematics School, I gained no advantage at the entrance examinations. A fair rule indeed! It was challenging, but we both succeeded, becoming first-year students at Novosibirsk State University.

I won’t delve into our student years here. While it was an exhilarating experience for us, from an outside perspective, things were probably quite ordinary. However, as we approached completing our M.A. theses and especially entering the Ph.D. study, we found ourselves in an entirely new reality.

Now, I am compelled to recount a sordid tale of antisemitic persecution in Akademgorodok, beginning in 1968 and culminating in 1971, the year of our graduation. I cherished my life in Akademgorodok and am grateful to the kind individuals (some of whose names I will mention below) who assisted Boris and me, as well as many others, in navigating through those difficult times. However, omitting this part of our lives would be impossible; it was crucially significant for me, and likely for Boris as well.

From 1965, when I arrived there, until 1968, I observed no signs of antisemitism in Akademgorodok. Perhaps it existed among the higher social echelons, but not among the students. The entrance examinations to Novosibirsk State University were utterly fair! While I cannot provide documented evidence, I believe that around 30 percent of the new mathematics students in 1966 were Jews from various parts of the USSR. They knew that in Novosibirsk, there was a fair chance! However, in 1968 and 1969, the situation underwent a dramatic shift. Some of our Russian student acquaintances now served on the new entrance examination committees, intentionally created to hinder the chances of Jewish applicants. Occasionally, they confided in us, maintaining the open traditions of our old friendship, about what transpired during these entrance examinations. They recounted the now well-known tales of exceptionally challenging mathematical problems posed to Jewish candidates, among other tactics.

We were also informed that the authorities' decision was to make Akademgorodok "judenrein" — free of Jews. I am uncertain whether this German term was used by the Akademgorodok Party Committee, but this is what was communicated to us. While all this was disconcerting, it did not directly affect us; we were veterans of 3–4 years, not the unfortunate new entrants.

However, in the spring of 1971, the year of our graduation, our turn arrived. Suddenly (for us, as we had not taken the earlier warning signs seriously), two-thirds of our top Jewish graduates received insufficient grades in their final exams. They could no longer aspire to continue their doctoral studies at Novosibirsk State University, or anywhere in Akademgorodok. This was a devastating blow! Both Boris and I weathered this challenging experience successfully (mostly due to the efforts of our mentors)! We could carry on! And at that moment, I still did not fully comprehend what was happening! This purge expelled our best and closest friends, and I merely participated in their farewell graduation celebrations, still hopeful for a bright future.

Allow me to digress briefly about myself — it's a rather amusing anecdote! Thanks to the vigorous efforts of my advisor, Vladimir Ivanovich Kuzminov, I was to get a starting research position (as a so-called "stager") at the Institute of Mathematics. This position was deemed secure: theoretically it could withstand even mediocre grades in the final exams. However, I harbored no illusions — I was certain they could find fault even with the stager position. Indeed, in April 1971, I was promptly selected for immediate military service, a fate usually reserved before 1971 for relatively weak graduates. This time, the list of potential servicemen included 13 highly accomplished graduates, among them 10 Jews, including myself (but not Boris), and 3 Russians. The absence of Boris's name from this glorious list may be understood from what is explained below: the main target was not him but his advisor Taitslin.

Naturally, my stager position at the Institute of Mathematics was revoked. We, the new servicemen, were slated to serve in the Moscow rocket defense. If I indeed entered this service, I could not entertain the notion of emigrating to Israel for at least 20 years, due to the secrecy restrictions. Moreover, even *perestroika*, as we now know, would not have provided much respite: mass emigration to Israel commenced only in 1991, precisely 20 years after 1971.

Eventually, after some introspection, I wholeheartedly accepted this shift in my fate. I commenced a series of farewell gatherings. Surprisingly, I quickly discovered that this was a far simpler existence than pursuing mathematics or preparing for Ph.D. entrance exams. Particularly since we, the Jewish candidates, were uncertain where we might face persecution: in mathematics exams, in Communist philosophy, or elsewhere.

Thankfully, Marshal Grechko, the defense minister of the USSR, struck out all the Jewish names from the Novosibirsk list of potential servicemen. Marshal Grechko surely had to personally intervene in this minor issue solely due to the global significance of the Moscow rocket defense.

Now I returned to my tribulations, and Boris and I found ourselves facing the oral Ph.D. entrance exams. The first was in mathematics. Remarkably, the math exams were conducted rather transparently: all the Jews who had survived the purge at the graduation exams (around 70 such Jews, as I recall, from different faculties of Novosibirsk State University) were assigned to a single examination committee. Its head was Academician Yanenko, and after the first day of exams, we all knew precisely what was happening. Typically, Yanenko would interrupt the examinee after three minutes, declaring, “No, this is not it!” If the examinee persisted, Yanenko would repeat this scenario more frequently and forcefully until, in 8–10 minutes, he rendered the verdict: “You may know something, but at best to the level of 3.” A grade 3 in mathematics did not qualify for doctoral studies.

This scheme operated flawlessly. As far as I am aware, only three Jews out of the 70 managed to breach this absolute defense: Boris, myself, and Grisha Soifer. My success story was brief and straightforward: during my examination, at the fourth minute, Academician Yanenko abruptly departed — perhaps for a place even academicians have to reach on foot. The rest of the committee promptly awarded me the highest grade, 5, and then, with great insistence, urged me to leave the room. I complied most happily. I am uncertain how Grisha Soifer achieved it! As for Boris, it seems his admission was a move in a campaign against his unofficial advisor, Michael Abramovich Taitslin. Well before our doctoral study entrance exams described above, Taitslin had apparently been warned that he had more than enough Jewish students and could not accept any more. Now, for Boris, the pivotal events occurred on the last possible day to apply to the Ph.D. entrance exams. Among the required application documents was a written agreement from the prospective advisor. In Boris’s case, his official advisor was Yu. L. Ershov. Ershov was expected to agree, but on this final day he let Boris know that he could not serve as his advisor due to several significant administrative and scientific reasons. Boris's only hope of continuing his Ph.D. studies in Akademgorodok was if Taitslin agreed to be his advisor in the remaining few hours. And he did agree, despite the warnings. Boris passed Yanenko's scrutiny without difficulty, but as a result of his generosity in taking on Boris, Taitslin was expelled from the Institute of Mathematics and later from Akademgorodok altogether.

The remaining Ph.D. entrance exams, including Communist philosophy, proceeded smoothly. And now, after all these trials, there began a joyous season for me and, I presume, for Boris. Three years of unfettered scientific research and study, with scarcely a thought of the entrance battles. This period was profoundly significant for all our future endeavors! I am grateful to the circumstances for this felicitous interlude.

I wish to express my gratitude to another outstanding individual who greatly aided many of us during difficult times: Alexei Andreevich Lyapunov. Sometime during my three-year Ph.D. study, the Novosibirsk Energy Institute, where I was slated to work upon completing my Ph.D., announced that they were no longer interested in my services. Presumably, this was due to the burgeoning phenomenon of Jewish emigration to Israel, and they wished to avoid potential entanglements with me. I was apprehensive that my Ph.D. study might be affected, but nothing of the sort occurred. A few weeks after the announcement from the Novosibirsk Energy Institute, a young and athletic-looking individual knocked on my door. He relayed that he had been tasked by Lyapunov, one of the prominent scientists in Akademgorodok, to contact me. Lyapunov's message to me was that he might wield some influence at the Novosibirsk Energy Institute and, if I were interested, he could exert his best efforts to reinstate my position. I was deeply appreciative but requested him not to intervene. By then, I was already seriously contemplating emigrating to Israel.

I previously mentioned my advisor Vladimir Ivanovich Kuzminov. Allow me to also acknowledge Igor Aleksandrovich Shvedov. During those trying times, they both did what they could to assist their Jewish students. And, of course, once again, I express my gratitude to Alexei Andreevich Lyapunov.

One more pleasant recollection from those halcyon days. One autumn (probably, of 1973) Boris and I both served as group leaders in the obligatory autumn student agricultural service — in our case, this involved harvesting potatoes. My group consistently ranked last in the daily ratings, and I felt dejected by my failure. However, a couple of years later, I was appointed group leader of the combined group of female students from the Tashkent Polytechnical Institute. This time, the obligatory autumn student agricultural service involved picking cotton. Again, my group languished at the bottom of the daily ratings, but this time, I was not disheartened by my failure. As far as I know, Boris missed out on this opportunity to lead students in picking cotton.

In 1974, I left Novosibirsk for Barnaul, and after three months there, I moved to Tashkent. Finally, in 1978, I immigrated to Israel. Boris relocated to Kemerovo, and we only crossed paths again in the late 1990s. Since then, we have been meeting more or less regularly.

References

[1] G. Binyamini and D. Novikov, "Wilkie's conjecture for restricted elementary functions", *Ann. of Math.* (2) **186**:1 (2017), 237–275. MR Zbl

[2] G. Binyamini and D. Novikov, "Complex cellular structures", *Ann. of Math.* (2) **190**:1 (2019), 145–248. MR Zbl

[3] G. Binyamini and D. Novikov, "The Yomdin–Gromov algebraic lemma revisited", *Arnold Math. J.* **7**:3 (2021), 419–430. MR Zbl

[4] G. Binyamini, H. Schmidt, and A. Yafaev, "Lower bounds for Galois orbits of special points on Shimura varieties: a point-counting approach", *Math. Ann.* **385**:1-2 (2023), 961–973. MR Zbl

[5] G. Binyamini, D. Novikov, and B. Zak, "Wilkie's conjecture for Pfaffian structures", *Ann. of Math.* (2) **199**:2 (2024), 795–821. MR Zbl

[6] E. Bombieri and J. Pila, "The number of integral points on arcs and ovals", *Duke Math. J.* **59**:2 (1989), 337–357. MR Zbl

[7] E. Bombieri, D. Masser, and U. Zannier, "Anomalous subvarieties—structure theorems and applications", *Int. Math. Res. Not.* **2007**:19 (2007), art. id. rnm057. MR Zbl

[8] D. Burguet, "A proof of Yomdin–Gromov's algebraic lemma", *Israel J. Math.* **168** (2008), 291–316. MR Zbl

[9] D. Burguet, "SRB measures for C^∞ surface diffeomorphisms", *Invent. Math.* **235**:3 (2024), 1019–1062. MR Zbl

[10] D. Burguet, G. Liao, and J. Yang, "Asymptotic h-expansiveness rate of C^∞ maps", *Proc. Lond. Math. Soc.* (3) **111**:2 (2015), 381–419. MR Zbl

[11] R. Cluckers, O. Friedland, and Y. Yomdin, "Doubling coverings via resolution of singularities and preparation", *Commun. Contemp. Math.* **23**:2 (2021), art. id. 2050018. MR Zbl

[12] O. Friedland and Y. Yomdin, "Doubling coverings of algebraic hypersurfaces", *Pure Appl. Funct. Anal.* **2**:2 (2017), 221–241. MR Zbl

[13] O. Friedland and Y. Yomdin, "Doubling chains on complements of algebraic hypersurfaces", *J. Anal. Math.* **140**:2 (2020), 617–636. MR Zbl

[14] M. Gromov, "Entropy, homology and semialgebraic geometry (after Y. Yomdin)", pp. exposé 663, pp. 225–240 in *Séminaire Bourbaki*: *vol.* 1985/86, Astérisque **145–146**, Soc. math. de France, 1987.

[15] D. Haviv and Y. Yomdin, "Uniform approximation of near-singular surfaces", *Theoret. Comput. Sci.* **392**:1-3 (2008), 92–100. MR Zbl

[16] O. Marmon, "A generalization of the Bombieri–Pila determinant method", *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **377** (2010), 63–77, 242. In Russian; translation in *J. Math. Sci.* (NY) **171**:6 (2010), 736–744. MR Zbl

[17] J. Pila, *Point-counting and the Zilber–Pink conjecture*, Cambridge Tracts in Mathematics **228**, Cambridge University Press, 2022. MR Zbl

[18] J. Pila and A. J. Wilkie, "The rational points of a definable set", *Duke Math. J.* **133**:3 (2006), 591–616. MR Zbl

[19] J. Pila, A. N. Shankar, J. Tsimerman, H. Esnault, and M. Groechenig, "Canonical heights on Shimura varieties and the André–Oort Conjecture", preprint, 2022. Zbl arXiv 2109.08788v3

[20] R. Pink, "A combination of the conjectures of Mordell–Lang and André–Oort", pp. 251–282 in *Geometric methods in algebra and number theory*, Progr. Math. **235**, Birkhäuser, Boston, 2005. MR Zbl

[21] T. Scanlon, "A proof of the André-Oort conjecture via mathematical logic [after Pila, Wilkie and Zannier]", pp. exposé 1037, pp. 299–315 in *Séminaire Bourbaki*: *vol.* 2010/2011, Astérisque **348**, Soc. Math. de France, 2012. MR Zbl

[22] Y. Yomdin, "C^k-resolution of semialgebraic mappings: addendum to 'Volume growth and entropy' ", *Israel J. Math.* **57**:3 (1987), 301–317. MR Zbl

[23] Y. Yomdin, "Volume growth and entropy", *Israel J. Math.* **57**:3 (1987), 285–300. MR Zbl

[24] Y. Yomdin, "Local complexity growth for iterations of real analytic mappings and semicontinuity moduli of the entropy", *Ergodic Theory Dynam. Systems* **11**:3 (1991), 583–602. MR Zbl

[25] Y. Yomdin, "Analytic reparametrization of semi-algebraic sets", *J. Complexity* **24**:1 (2008), 54–76. MR Zbl

[26] Y. Yomdin, "Smooth parametrizations in dynamics, analysis, diophantine and computational geometry", *Jpn. J. Ind. Appl. Math.* **32**:2 (2015), 411–435. MR Zbl

[27] U. Zannier, *Some problems of unlikely intersections in arithmetic and geometry*, Annals of Mathematics Studies **181**, Princeton University Press, 2012. MR Zbl

[28] B. Zilber, "Exponential sums equations and the Schanuel conjecture", *J. London Math. Soc.* (2) **65**:1 (2002), 27–44. MR Zbl

Received 26 Apr 2024. Revised 1 Jul 2024.

YOSEF YOMDIN:

yosef.yomdin@weizmann.ac.il
Department of Mathematics, The Weizmann Institute of Science, 76100 Rehovot, Israel

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.599

The existential closedness and Zilber–Pink conjectures

Vahagn Aslanyan

Dedicated to Boris Zilber on the occasion of his 75th birthday.

We survey the history of, and recent developments on, two major conjectures originating in Zilber's model-theoretic work on complex exponentiation: existential closedness and Zilber–Pink. The main focus is on the modular versions of these conjectures and specifically on novel variants incorporating the derivatives of modular functions. The functional analogues of all the conjectures are already theorems, which we also present. The paper also contains some new results and conjectures.

1. Introduction

In the early 2000s, Boris Zilber [2002; 2005; 2015] produced an influential body of work around the model theory of the complex exponential field $\mathbb{C}_{\exp} := (\mathbb{C}; +, \cdot, \exp)$, where $\exp : z \to e^z$ is the exponential function. He showed that *Schanuel's conjecture* (SC for short) on the transcendence properties of exp (see Section 2A) plays a central role in the model-theoretic properties of $\mathbb{C}_{\exp}$. However, the conjecture is out of reach — it implies the algebraic independence of e and π over the rationals, which is a long-standing unsolved problem. This makes it hard to understand the model theory of $\mathbb{C}_{\exp}$. So Zilber constructed algebraically closed fields of characteristic 0 equipped with a unary function, which satisfies some of the basic properties of $\mathbb{C}_{\exp}$ and, most importantly, (the analogue of) Schanuel's conjecture. He then isolated and axiomatised the "most" existentially closed ones among these exponential fields by a Hrushovski style amalgamation-with-predimension construction. These are called *pseudo-exponential fields*. While these models are not existentially closed in the first-order sense, they are existentially closed in certain "tame" extensions. The axiom guaranteeing this is known as *strong existential closedness* or *strong exponential closedness*, or SEC for short.

Supported by Leverhulme Trust Early Career Fellowship ECF-2022-082 (at the University of Leeds) and partially supported by EPSRC Open Fellowship EP/X009823/1 and DKO Fellowship (at the University of Manchester).
MSC2020: primary 11F03, 11G18; secondary 03C60, 12H05, 12L12.
Keywords: j-function, Schanuel's conjecture, Zilber–Pink, existential closedness.

Zilber showed that his axiomatisation of pseudo-exponential fields is uncountably categorical. In particular, there is a unique pseudo-exponential field of cardinality of the continuum, denoted by $\mathbb{B}_{\exp}$. Zilber conjectured that $\mathbb{B}_{\exp}$ is isomorphic to $\mathbb{C}_{\exp}$. This is equivalent to the combination of two conjectures — Schanuel's conjecture and the strong exponential closedness conjecture stating that SEC holds on $\mathbb{C}_{\exp}$. A variant of the SEC conjecture, known as *exponential closedness* or *existential closedness* (EC for short) is currently an active research field in model theory. It states roughly that systems of equations involving field operations and the complex exponential function have solutions unless they are "overdetermined" (i.e., the number of independent equations is larger than the number of variables). The notion of overdetermined systems is in fact related to Schanuel's conjecture: a system is overdetermined if its solution would be a counterexample to Schanuel's conjecture. As the name suggests, SEC is a strong version of EC guaranteeing that under certain conditions, systems of exponential equations have generic solutions.

Zilber's work on the model theory of complex exponentiation also gave rise to a Diophantine conjecture: the *conjecture on intersections with tori*, or CIT for short. It states roughly that intersections of algebraic varieties with torsion cosets of algebraic tori, whose dimensions are larger than expected, are governed by finitely many torsion cosets of algebraic tori. The statement makes sense in the more general setting of semiabelian varieties which gives rise to the *conjecture on intersections with semiabelian varieties*, or CIS for short. Both CIT and CIS were proposed in [Zilber 2002] and independently by Bombieri, Masser, and Zannier in [Bombieri et al. 2007]. The Manin–Mumford and Mordell–Lang conjectures are special cases of CIS. Zilber used CIT to deduce a uniform version of Schanuel's conjecture from itself, which then was used to establish some partial results towards exponential closedness (see [Zilber 2002]).

SC, EC, and CIT are quite general in form; replacing the exponential function by other transcendental functions often allows one to formulate analogues of these conjectures in other settings. Most notably, such analogues have been extensively explored for modular functions and, in particular, the j-invariant. However, these analogues are being studied for other reasons too: the modular analogue of Schanuel's conjecture is a special case of the Grothendieck–André generalised period conjecture (see [Bertolin 2002, 1.3 Corollaire; André 2004, §23.4.4; Aslanyan et al. 2023a, §6.3]), EC in that setting is a natural problem in complex geometry and model theory, and the analogue of CIT is a special case of the Zilber–Pink conjecture for (mixed) Shimura varieties, henceforth referred to as ZP. The latter was proposed by Pink (independently from Zilber and Bombieri, Masser, and Zannier) as a far-reaching conjecture unifying the André–Oort, André–Pink–Zannier, Manin–Mumford, and Mordell–Lang conjectures [Pink 2005a; 2005b].

Furthermore, from a model-theoretic point of view, if both SC and EC hold then in a certain sense they give a "complete" list of properties (non-first-order axioms) of the function under consideration.[1] This is formalised by Zilber's categoricity and quasiminimality theorem in the exponential setting. There is no such theorem in the modular setting and there cannot be one, for the upper half-plane (hence the set of the reals) is definable from the graph of j, but the philosophy of SC and EC together giving a full description of the algebraic and transcendental properties of j still applies. It is likely that a formal categoricity/quasiminimality result can be established for some relations defined in terms of j (which give proper reducts of the complex field with j); this is part of our current research programme.

In this paper we present the above-mentioned conjectures in the exponential and modular settings, mostly focusing on the latter. As pointed out above, the modular variants of these conjectures are in part motivated by their exponential counterparts. However, there are some inherent differences between the two settings resulting in quite different methods and approaches, although some methods work in both contexts. One such difference is that unlike exponential functions, which are defined on the whole complex plane, modular functions are defined only on the upper half-plane. These spaces are "geometrically different", which accounts for different approaches to EC and ZP in these two settings. This also makes the model-theoretic treatment of modular functions significantly harder. For example, direct counterparts of many aspects of Zilber's work on exponentiation, e.g., categoricity and quasiminimality, fail gravely in the modular setting (as explained above).

Further, modular functions satisfy third-order differential equations as opposed to first-order differential equations for exponential functions. So we can consider SC, EC, and ZP for modular functions together with their first two derivatives (the third one being algebraic over these). This generalisation makes the problems more challenging, but it also gives a deeper insight into them by providing a broader model-theoretic picture. Let us briefly discuss two more reasons to consider SC, EC, and ZP for modular functions together with derivatives. Often when dealing with variants of these conjectures, not least their differential versions, even when derivatives are not considered, the approaches and techniques require looking at the derivatives anyway (see, for instance, [Aslanyan et al. 2021; Aslanyan 2022b]). Also, modular forms of weight 2 are the derivatives of modular functions (weight 0), which means that studying these problems for modular forms of weight 2 (without derivatives) is the same as studying them for the first derivatives of modular functions.

We state several versions of the conjectures in this new setting, some of which have appeared in the literature while others are new. We then explain the relationship

[1]This means, in particular, that if EC holds then SC is the strongest possible transcendence statement about the function under consideration.

between these various conjectures and present their functional variants, all of which were proven in recent years, save for Ax's original theorem proven in 1971.

1A. ***Abbreviations.*** In the paper we consider several variants of three conjectures: Schanuel's conjecture, the existential closedness conjecture, and the Zilber–Pink conjecture. We use abbreviations to refer to those conjectures, and for the convenience of the reader we list some of these abbreviations below.

Schanuel

- SC — Schanuel conjecture
- MSC — modular Schanuel conjecture
- MSCD — modular Schanuel conjecture with derivatives

Existential closedness

- EC — existential closedness or exponential closedness
- MEC — modular existential closedness
- MECD — modular existential closedness with derivatives

Zilber–Pink

- CIT — conjecture on intersections with tori
- ZP — Zilber–Pink
- MZP — modular Zilber–Pink
- MZPD — modular Zilber–Pink with derivatives

1B. ***Dedication.*** This paper is dedicated to Boris Zilber on the occasion of his 75th birthday, and is motivated by his work. Boris was my DPhil supervisor (jointly with Jonathan Pila) at the University of Oxford from 2013 to 2017. His guidance has been instrumental in shaping my mathematical thinking and research interests, and his continued support, both throughout my DPhil and after that, has been tremendously helpful in my mathematical career. The hours spent with Boris at the Mathematical Institute and at Merton are some of my fondest memories of Oxford. I would like to thank him for everything and wish him a happy 75th birthday.

2. The exponential setting

In this section we look briefly at Zilber's work on model theory of complex exponentiation and the conjectures it gave rise to.

2A. ***Schanuel's conjecture and exponential closedness.*** We begin by formulating Schanuel's conjecture.

Conjecture 2.1 (Schanuel's conjecture — SC [Lang 1966, p. 30]). *For any $\mathbb{Q}$-linearly independent complex numbers $z_1, \ldots, z_n$,*

$$\operatorname{td}_{\mathbb{Q}} \mathbb{Q}(z_1, \ldots, z_n, e^{z_1}, \ldots, e^{z_n}) \geq n,$$

where td *stands for transcendence degree.*

This conjecture is believed to capture *all* transcendence properties of the exponential function. This can and will shortly be explained in a more precise sense. For now let us mention that Schanuel's conjecture for $n = 2$ already implies the algebraic independence of e and π by choosing $z_1 = \pi i$, $z_2 = 1$, which is a long-standing open problem. Thus, even for $n = 2$ the conjecture is out of reach of current methods. However, partial results towards this conjecture are known, including the Lindemann–Weierstrass theorem and the Gelfond–Schneider theorem.

Zilber [2005] presented a novel model-theoretic approach to Schanuel's conjecture. He constructed algebraically closed fields of characteristic 0 equipped with a unary function, known as *pseudo-exponentiation*, satisfying certain basic properties of the complex exponential functions and some desirable properties, not least the analogue of Schanuel's conjecture. He axiomatised these structures in the language $L_{\omega_1,\omega}(Q)$, where Q is a quantifier for "there are uncountably many", and showed that the resulting theory is categorical in uncountable cardinals. The unique model of cardinality $2^{\aleph_0}$ is called *the pseudo-exponential field* or *the Zilber field* and is usually denoted by $\mathbb{B}_{\exp}$. Zilber then conjectured that $\mathbb{B}_{\exp}$ is isomorphic to $\mathbb{C}_{\exp}$. This shows, in a sense, that Schanuel's conjecture must play a central role in the model theory of $\mathbb{C}_{\exp}$.

Zilber verified that all of the axioms of pseudo-exponentiation hold in $\mathbb{C}_{\exp}$ save for Schanuel's conjecture and an axiom called *strong exponential closedness* (SEC). So Zilber's conjecture that $\mathbb{B}_{\exp} \cong \mathbb{C}_{\exp}$ is equivalent to the conjunction of Schanuel's conjecture and the strong exponential closedness conjecture (stating that the axiom holds in $\mathbb{C}_{\exp}$).

Let us explain what (strong) exponential closedness means. Schanuel's conjecture can be interpreted as a statement about nonsolvability of certain systems of equations, which we demonstrate on an example below.

Example 2.2 [Aslanyan et al. 2023b]. Assume e and π are algebraically independent over $\mathbb{Q}$. Then for any nonconstant polynomial $p(X, Y) \in \mathbb{Q}[X, Y]$ the system $e^z = 1$, $p(z, e) = 0$ does not have solutions in $\mathbb{C}$. On the other hand, if e and π are algebraically dependent, then for some p that system does have a complex solution.

Another reason for a system not to have a solution is when the system is incompatible with the identity $e^{x+y} = e^x e^y$.

Example 2.3. The system $z_2 = z_1 + 1$, $3e^{z_1} = e^{z_2}$ does not have a solution, for the first equation implies $e^{z_2} = e \cdot e^{z_1}$ and $e \neq 3$. On the other hand, the system

$z_2 = z_1 + 1$, $e^{z_2} = z_1$, $e^{z_2} = e \cdot e^{z_1}$ does have solutions even though there are three equations in two variables. Of course, the three equations are not "analytically" independent — the third one follows from the first one by taking exponentials of both sides — but they are algebraically independent.

In general, systems incompatible with the functional equation of exp are not solvable. Moreover, SC implies that if a system is "overdetermined", e.g., n variables with more than n algebraically independent equations, then there is no solution, unless the system can somehow be reduced using the functional equation $e^{x+y} = e^x e^y$. With this interpretation SC becomes more natural, and exponential closedness (EC) is its dual conjecture stating roughly that a system of exponential equations does have a solution in $\mathbb{C}$ unless having a solution contradicts Schanuel's conjecture. Let us give a precise statement in geometric terms, observing first that understanding the solvability of systems of exponential equations is equivalent to understanding when algebraic varieties contain exponential points (i.e., points of the form $(\bar{z}, \exp(\bar{z}))$). For instance, the equation $e^{e^z} + z - 1 = 0$ has a solution if and only if the variety $V \subseteq \mathbb{C}^2 \times (\mathbb{C}^\times)^2$ (with coordinates (x_1, x_2, y_1, y_2) defined by the equations $x_2 = y_1$, $y_2 + x_1 - 1 = 0$) contains an exponential point.

Conjecture 2.4 (exponential closedness — EC [Zilber 2005; Bays and Kirby 2018]). *Let $V \subseteq \mathbb{C}^n \times (\mathbb{C}^\times)^n$ be a free and rotund variety. Then V contains a point of the form $(z_1, \dots, z_n, e^{z_1}, \dots, e^{z_n})$.*

Freeness and rotundity are the conditions that make sure containing an exponential point does not contradict SC, as illustrated on the above examples. Now we define these notions precisely.

Definition 2.5. An irreducible variety $V \subseteq \mathbb{C}^n \times (\mathbb{C}^\times)^n$ is *additively* (resp. *multiplicatively*) *free* if its projection to $\mathbb{C}^n$ (resp. $(\mathbb{C}^\times)^n$) is not contained in a translate of a $\mathbb{Q}$-linear subspace of $\mathbb{C}^n$ (resp. algebraic subgroup of $(\mathbb{C}^\times)^n$). A variety is called *free* if it is additively and multiplicatively free.

We let $\bar{x}$ and $\bar{y}$ denote the coordinates on $\mathbb{C}^n$ and $(\mathbb{C}^\times)^n$, respectively. For a $k \times n$ matrix M of integers we define $[M] : \mathbb{C}^n \times (\mathbb{C}^\times)^n \to \mathbb{C}^k \times (\mathbb{C}^\times)^k$ to be the map given by $[M] : (\bar{x}, \bar{y}) \mapsto (M\bar{x}, \bar{y}^M)$, where

$$(M\bar{x})_i = \sum_{j=1}^n m_{ij} x_j \quad \text{and} \quad (\bar{y}^M)_i = \prod_{j=1}^n y_j^{m_{ij}}.$$

Definition 2.6. An irreducible variety $V \subseteq \mathbb{C}^n \times (\mathbb{C}^\times)^n$ is *rotund* if for any $1 \le k \le n$ and any $k \times n$ matrix M of integers $\dim [M](V) \ge \operatorname{rk} M$.

Since exp maps $\mathbb{Q}$-linear equations to multiplicative ones, if the projections of V satisfy either a linear or multiplicative equation and we want it to contain an exponential point, then these equations should match; otherwise they would not be compatible with exp. Freeness takes care of this scenario by ensuring no such equations hold on the variety. Rotundity comes from SC; it states that V and its various projections given by the maps $[M]$ have sufficiently large dimension so an exponential point in V would not give a counterexample to SC.

Now we can formulate SEC, which is a strong version of EC.

Conjecture 2.7 (strong exponential closedness — SEC [Zilber 2005; Bays and Kirby 2018]). *Let $V \subseteq \mathbb{C}^n \times (\mathbb{C}^\times)^n$ be a free and rotund variety. Then for every finitely generated field $K \subseteq \mathbb{C}$ over which V is defined, there is a point*

$$(z_1, \dots, z_n, e^{z_1}, \dots, e^{z_n}) \in V$$

which is generic in V over K, that is, $\operatorname{td}_K K(\bar{z}, e^{\bar{z}}) = \dim V$.

It is obvious that SEC implies EC. The converse is also true assuming SC and CIT hold (see [Eterović 2022; Kirby and Zilber 2014]).

Remark 2.8. The Rabinowitsch trick can be used to show that EC implies that a free and rotund variety contains a Zariski dense set of exponential points (see [Kirby 2009, Theorem 4.11] and [Aslanyan 2022a, Proposition 4.34]), but a priori such a set may not contain a generic point.

2B. ***Conjecture on intersections with tori.*** Zilber [2002] studied the solvability of exponential sums equations as a special case of the exponential closedness conjecture. In order to prove that certain systems of such equations are solvable, he needed a uniform version of Schanuel's conjecture. He then proposed a Diophantine conjecture, called the *conjecture on intersections with tori*, or CIT for short, which acts as the difference between SC and uniform SC. The conjecture states roughly that when we intersect an algebraic variety with algebraic tori then we do not expect to get too many intersections which are *atypically large*. We will shortly give a precise formulation, but we need to introduce some notions first.

Let V and W be subvarieties of some variety S. A nonempty component X of the intersection $V \cap W$ is *atypical* in S if $\dim X > \dim V + \dim W - \dim S$, and *typical* if $\dim X = \dim V + \dim W - \dim S$. Note that if S is smooth then a nonstrict inequality always holds.

An *algebraic torus* is an irreducible algebraic subgroup of a multiplicative group $(\mathbb{C}^\times)^n$. Algebraic subgroups of $(\mathbb{C}^\times)^n$ (not necessarily irreducible) are defined by multiplicative equations of the form $y_1^{m_1} \cdots y_n^{m_n} = 1$ with $m_i \in \mathbb{Z}$. Any system of such equations (if consistent) defines an algebraic group. It splits as the union of an algebraic torus (the component containing the identity) and its finitely many translates by torsion points. Torsion cosets of algebraic tori are called *special* varieties. For an algebraic variety $V \subseteq (\mathbb{C}^\times)^n$ an *atypical subvariety* of V is an atypical component of an intersection of V with a special variety T.

Now we are ready to formulate the conjecture on intersections with tori, which is the Zilber–Pink conjecture for algebraic tori. There are many equivalent forms of the conjecture; we consider four of them.

Conjecture 2.9 (conjecture on intersections with tori — CIT [Zilber 2002; Bombieri et al. 2007; Pila 2022]). *Let $V \subseteq (\mathbb{C}^\times)^n$ be an algebraic variety.*

(1) *There is a finite collection Σ of proper special subvarieties of $(\mathbb{C}^\times)^n$ such that every atypical subvariety of V is contained in some $T \in \Sigma$.*

(2) *V contains only finitely many maximal atypical subvarieties.*

(3) *Let $\mathrm{Atyp}(V)$ be the union of all atypical subvarieties of V. Then $\mathrm{Atyp}(V)$ is contained in a finite union of proper special subvarieties of $(\mathbb{C}^\times)^n$.*

(4) *$\mathrm{Atyp}(V)$ is a Zariski closed subset of V.*

2C. *Functional/differential variants.* We have so far considered three conjectures for $(\mathbb{C}^\times)^n$, namely, SC, EC, and CIT. As pointed out above, Schanuel's conjecture is out of reach, CIT is wide open, and while EC is more tractable, it is also open. In spite of that, the functional analogues of all three conjectures are known.

Ax proved a functional analogue of Schanuel's conjecture in 1971. Below in a differential field $(F; +, \cdot, D_1, \dots, D_m)$ we define a relation $\mathrm{Exp}(\bar{x}, \bar{y})$ as the set of all $(\bar{x}, \bar{y}) \in F^n \times (F^\times)^n$ for which $D_k y_i = y_i D_k x_i$ for all k, i. Then $\mathrm{Exp}(F)$ is the set of all tuples $(\bar{x}, \bar{y}) \in F^n \times (F^\times)^n$ with $F \models \mathrm{Exp}(\bar{x}, \bar{y})$ (for all n).[2]

Theorem 2.10 (Ax–Schanuel [Ax 1971, Theorem 3]). *Let $(F; +, \cdot, D_1, \dots, D_m)$ be a differential field with field of constants $C = \bigcap_{k=1}^m \ker D_k$. Let also $(x_i, y_i) \in F^2$, $i = 1, \dots, n$, be such that $(\bar{x}, \bar{y}) \in \mathrm{Exp}(F)$. Assume $x_1, \dots, x_n$ are $\mathbb{Q}$-linearly independent mod C, that is, they are $\mathbb{Q}$-linearly independent in the quotient vector space F/C. Then $\mathrm{td}_C\, C(\bar{x}, \bar{y}) \geq n + \mathrm{rk}(D_k x_i)_{i,k}$.*

Ax's proof of this theorem is differential algebraic. There is an equivalent complex analytic formulation of Ax–Schanuel (the equivalence follows from Seidenberg's embedding theorem). Tsimerman [2015] gave a new proof of that complex analytic statement based on o-minimality.

The differential version of EC for fields with several commuting derivations was established recently by Aslanyan, Eterović, and Kirby.

Theorem 2.11 (differential EC [Aslanyan et al. 2021, Theorem 4.3]). *For a differential field $(F; +, \cdot, D_1, \dots, D_m)$ with m commuting derivations, let $V \subseteq F^{2n}$ be a rotund variety. Then there exists a differential field extension K of F such that $V(K) \cap \mathrm{Exp}(K) \neq \varnothing$. In particular, when F is differentially closed,*

$$V(F) \cap \mathrm{Exp}(F) \neq \varnothing.$$

[2]More generally, here and later, given a relation R on a structure M, we write $R(M)$ for the set of all tuples from M satisfying the relation R.

The proof of this theorem uses some important differential algebraic ideas from [Kirby 2009], where the case of ordinary differential fields was treated. Kirby's approach (which in fact contains some inaccuracies and is not complete) is based on Ax's proof of the Ax–Schanuel theorem, while the argument given in [Aslanyan et al. 2021] uses the statement of Ax–Schanuel as a black box and works quite generally.

Example 2.12. In the above theorem the variety V need not be free. However, freeness is a necessary condition in EC. For example, the variety $V \subseteq \mathbb{C}^2 \times (\mathbb{C}^\times)^2$ defined by the equations $x_2 = x_1$, $y_2 = y_1 + 1$, which is rotund but not free, cannot intersect the graph of any function. But it does intersect $\mathrm{Exp}(K)$ for any differential field K — indeed any constant point in V is actually in $\mathrm{Exp}(K)$.

Finally, the following functional analogue of CIT was established independently in [Zilber 2002] and in [Bombieri et al. 2007]. Both proofs rely on the Ax–Schanuel theorem. Kirby [2009] adapted Zilber's argument and gave a new proof using the uniform version of Ax–Schanuel, which follows from Ax–Schanuel by an application of the compactness theorem of first-order logic (see [Kirby 2009, Theorem 4.3]).

Theorem 2.13 (functional CIT [Zilber 2002; Bombieri et al. 2007; Kirby 2009]). *For every subvariety $V \subseteq (\mathbb{C}^\times)^n$ there is a finite collection Σ of proper subtori of $(\mathbb{C}^\times)^n$ such that every atypical component of an intersection of V with a coset of a torus is contained in a coset of some torus $T \in \Sigma$.*

Theorem 2.13 is indeed a functional version of CIT as it talks about *weakly special varieties* (arbitrary cosets of tori) and positive-dimensional atypical intersections. In other words, it can be thought of as CIT over function fields, where we work modulo the constants (in this case, the field of complex numbers). It does not say anything about special points or special coordinates in atypical intersections, so it is often called the geometric component of CIT (i.e., CIT without its arithmetic component). Since its statement is algebraic (rather than differential algebraic), it is also often called weak CIT although, strictly speaking, it is not a weak version of CIT.

In addition to the above-mentioned theorems, some other partial results have also been obtained towards EC and CIT in recent years. For EC see [Aslanyan et al. 2023b; Gallinaro 2021; 2023; Brownawell and Masser 2017; D'Aquino et al. 2018]. It would be impractical to try to give a comprehensive list of references for CIT and its generalisations to semiabelian varieties, so we refer the reader to Pila's recent book [2022] and references therein.

3. The modular setting

We let $\mathbb{H} \subseteq \mathbb{C}$ denote the complex upper half-plane and $j : \mathbb{H} \to Y(1)$ denote the modular j-function. We identify the modular curve $Y(1)$ with the complex affine line $\mathbb{C}$.

Recall that the j-function is invariant under the linear fractional action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$ and behaves nicely under the action of $\mathrm{GL}_2^+(\mathbb{Q})$ (where $+$ denotes positive determinant). More precisely, there is a collection of *modular polynomials* $\Phi_N(Y_1, Y_2) \in \mathbb{Z}[Y_1, Y_2]$, $N \in \mathbb{N}$, such that

$$\forall z_1, z_2 \in \mathbb{H}\ \big(\exists g \in \mathrm{GL}_2^+(\mathbb{Q}) \text{ with } z_2 = gz_1 \\ \text{iff } \exists N \in \mathbb{N} \text{ such that } \Phi_N(j(z_1), j(z_2)) = 0\big).$$

These correspondences are often referred to as the "functional equations" of the j-function. They are analogous to the functional equation $e^{x+y} = e^x e^y$ of the exponential function. This analogy allows one to state the modular counterparts of the exponential conjectures mentioned in the previous section, and that is what we do in this section. We focus on the j-function as other modular functions can be treated similarly, and often results about other modular functions can be deduced from those about j since j is a uniformiser for the modular group: it generates the field of all modular functions.

Now let us introduce some notation that will be used throughout the rest of the paper.

Notation. Let n be a positive integer, $k \leq n$ and $1 \leq i_1 < \cdots < i_k \leq n$.

- Subsets of $\mathbb{C}^{2n}$ (e.g., $\mathbb{H}^n \times \mathbb{C}^n$) are interpreted as subsets of $\mathbb{C}^n \times \mathbb{C}^n$, and we denote the coordinates on this space by $(\bar{x}, \bar{y})$.
- $\mathrm{Pr}_{\bar{x}} : \mathbb{C}^{2n} \to \mathbb{C}^n$ is the projection to the first n coordinates, and $\mathrm{Pr}_{\bar{y}} : \mathbb{C}^{2n} \to \mathbb{C}^n$ is the projection to the second n coordinates.
- $\mathrm{pr}_{\bar{\imath}} : \mathbb{C}^n \to \mathbb{C}^k$ is the map $\mathrm{pr}_{\bar{\imath}} : (x_1, \ldots, x_n) \mapsto (x_{i_1}, \ldots, x_{i_k})$.
- $\mathrm{Pr}_{\bar{\imath}} : \mathbb{C}^{2n} \to \mathbb{C}^{2k}$ denotes the map $\mathrm{Pr}_{\bar{\imath}} : (\bar{x}, \bar{y}) \mapsto (\mathrm{pr}_{\bar{\imath}}\, \bar{x}, \mathrm{pr}_{\bar{\imath}}\, \bar{y})$.
- By abuse of notation we let $j : \mathbb{H}^n \to \mathbb{C}^n$ denote all Cartesian powers of itself and $\Gamma_j \subseteq \mathbb{H}^n \times \mathbb{C}^n$ denote its graph.

3A. ***Modular Schanuel conjecture and modular existential closedness.*** We begin by stating the analogue of Schanuel's conjecture for the j-function. It is a special case of the Grothendieck–André generalised period conjecture [Bertolin 2002, 1.3 Corollaire; André 2004, §23.4.4; Aslanyan et al. 2023a, §6.3].

Conjecture 3.1 (modular Schanuel conjecture — MSC). *Let $z_1, \ldots, z_n \in \mathbb{H}$ be nonquadratic numbers with distinct $\mathrm{GL}_2^+(\mathbb{Q})$-orbits. Then*

$$\mathrm{td}_{\mathbb{Q}}\, \mathbb{Q}(z_1, \ldots, z_n, j(z_1), \ldots, j(z_n)) \geq n.$$

Schneider's theorem, stating that if both z and $j(z)$ are algebraic over $\mathbb{Q}$ then z must be a quadratic irrational number, is a special case of this conjecture.

As in the exponential setting, this conjecture can be interpreted as a statement about nonsolvability of certain systems of equations involving the j-function.

Example 3.2. Let $a, b \in \mathbb{Q}^{\text{alg}}$ be algebraic *nonspecial numbers*, that is, their preimages under j are not quadratic irrationals. By Schneider's theorem, these preimages cannot be algebraic. Consider the system

$$j(z_1) = a, \quad j(z_2) = b, \quad z_1^2 + z_2^2 + 1 = 0.$$

If this system has a solution, then $\operatorname{td}_{\mathbb{Q}} \mathbb{Q}(z_1, z_2, j(z_1), j(z_2)) = 1$. Hence, by MSC, either z_1 or z_2 must be a quadratic number or they must be in the same $\mathrm{GL}_2^+(\mathbb{Q})$-orbit. By our choice of a and b, the numbers z_1 and z_2 are transcendental over $\mathbb{Q}$, hence nonquadratic. If they satisfy a relation $z_2 = gz_1$ for some $g \in \mathrm{GL}_2^+(\mathbb{Q})$ then, together with the equation $z_1^2 + z_2^2 + 1 = 0$, we can conclude that $z_1, z_2 \in \mathbb{Q}^{\text{alg}}$, which is a contradiction. So MSC implies that the above system has no complex solutions. Note that it is overdetermined in the sense that we have 3 equations but only 2 variables.

Thus, we can propose a dual conjecture stating roughly that such a system always has a solution unless it contradicts MSC. We begin by recalling a few definitions from [Aslanyan 2022a].

Definition 3.3. Let $V \subseteq \mathbb{H}^n \times \mathbb{C}^n$ be an algebraic variety.

- V is Γ_j*-broad* if for any $1 \leq k_1 < \cdots < k_l \leq n$ we have $\dim \mathrm{Pr}_{\bar{k}} V \geq l$.
- V is *modularly free* if no equation of the form $\Phi_N(y_i, y_k) = 0$, or of the form $y_i = c$ with $c \in \mathbb{C}$ a constant, holds on V.
- V is $\mathrm{GL}_2^+(\mathbb{Q})$*-free* if no equation of the form $x_i = gx_k$ with $g \in \mathrm{GL}_2^+(\mathbb{Q})$, or of the form $x_k = c$ with $c \in \mathbb{H}$ a constant, holds on V.
- V is Γ_j*-free* if it is $\mathrm{GL}_2^+(\mathbb{Q})$-free and modularly free.
- V is Γ_j*-froad*[3] if it is Γ_j-free and Γ_j-broad.

Now we are ready to state the existential closedness conjecture.

Conjecture 3.4 (modular existential closedness — MEC [Aslanyan and Kirby 2022, Conjecture 1.2]). *Let $V \subseteq \mathbb{H}^n \times \mathbb{C}^n$ be an irreducible Γ_j-froad variety defined over $\mathbb{C}$. Then $V \cap \Gamma_j \neq \varnothing$.*

As in the exponential setting, we can consider a strong version of MEC — referred to as SMEC — stating that Γ_j-froad varieties contain generic points from Γ_j. Eterović [2022] proved that MSC, MZP (see below), MEC imply SMEC.

[3]To be pronounced like "fraud".

3B. *Modular Zilber–Pink.* Pink [2005a; 2005b] proposed a far-reaching conjecture in the setting of mixed Shimura varieties generalising the Manin–Mumford, Mordell–Lang, and André–Oort conjectures. That conjecture also generalises Zilber's CIT conjecture (although Pink came up with it independently from Zilber and Bombieri–Masser–Zannier) and is now known as the *Zilber–Pink conjecture*. Thus, CIT is the Zilber–Pink conjecture for algebraic tori. In this section we look at the Zilber–Pink conjecture in the modular setting, i.e., for $Y(1)^n$ (identified with $\mathbb{C}^n$ as usual).

Definition 3.5. • A j-*special* variety in $\mathbb{C}^n$ is an irreducible component of a variety defined by some modular equations $\Phi_N(y_k, y_l) = 0$.

• Let $V \subseteq \mathbb{C}^n$ be a variety. A j-*atypical* subvariety of V is an atypical component of an intersection $V \cap T$, where T is j-special.

As for CIT, modular Zilber–Pink has several equivalent formulations. Four of them are presented below.

Conjecture 3.6 (modular Zilber–Pink — MZP [Pila 2022, Conjecture 19.2]). *Let $V \subseteq \mathbb{C}^n$ be an algebraic variety. Let also* $\operatorname{Atyp}_j(V)$ *be the union of all j-atypical subvarieties of V. Then the following equivalent conditions hold.*

(1) *There is a finite collection Σ of proper j-special subvarieties of $\mathbb{C}^n$ such that every j-atypical subvariety of V is contained in some $T \in \Sigma$.*

(2) *V contains only finitely many maximal j-atypical subvarieties.*

(3) $\operatorname{Atyp}_j(V)$ *is contained in a finite union of proper j-special subvarieties of $\mathbb{C}^n$.*

(4) $\operatorname{Atyp}_j(V)$ *is a Zariski closed subset of V.*

As in the exponential setting, MZP and SC imply a uniform version of SC.

3C. *Functional/differential variants.* The j-function satisfies an order 3 algebraic differential equation over $\mathbb{Q}$, and none of lower order (see [Mahler 1969]). Namely, $\Psi(j, j', j'', j''') = 0$, where

$$\Psi(y_0, y_1, y_2, y_3) = \frac{y_3}{y_1} - \frac{3}{2}\left(\frac{y_2}{y_1}\right)^2 + \frac{y_0^2 - 1968y_0 + 2654208}{2y_0^2(y_0 - 1728)^2} \cdot y_1^2.$$

Thus

$$\Psi(y, y', y'', y''') = Sy + R(y)(y')^2,$$

where S denotes the *Schwarzian derivative* defined by $Sy = (y'''/y') - \frac{3}{2}(y''/y')^2$ and

$$R(y) = \frac{y^2 - 1968y + 2654208}{2y^2(y - 1728)^2}$$

is a rational function.

All functions $j(gz)$ with $g \in \mathrm{SL}_2(\mathbb{C})$ satisfy the equation $\Psi(y, y', y'', y''') = 0$ and all solutions (not necessarily defined on $\mathbb{H}$) are of that form (see [Freitag and Scanlon 2018, Lemma 4.2]).

Note that for nonconstant y, the relation $\Psi(y, y', y'', y''') = 0$ is equivalent to $y''' = \eta(y, y', y'')$, where

$$\eta(y, y', y'') := \frac{3}{2} \cdot \frac{(y'')^2}{y'} - R(y) \cdot (y')^3$$

is a rational function over $\mathbb{Q}$.

From now on, y', y'', y''' will denote some variables/coordinates and not the derivatives of y. Derivations of abstract differential fields will not be denoted by $'$. When we deal with actual functions though, $'$ will denote the derivative, e.g., j' is the derivative of j.

Definition 3.7. Let $(F; +, \cdot, D_1, \ldots, D_m)$ be a differential field with constant field $C = \bigcap_{k=1}^m \ker D_k$. We define a binary relation $\mathrm{D}\Gamma_j(x, y)$ by

$$\exists y', y'', y''' \Bigg[\Psi(y, y', y'', y''') = 0 \\ \wedge \bigwedge_{k=1}^{m} D_k y = y' D_k x \wedge D_k y' = y'' D_k x \wedge D_k y'' = y''' D_k x \Bigg].$$

The relation $\mathrm{D}\Gamma_j^{\times}(x, y)$ is defined by the formula $\mathrm{D}\Gamma_j(x, y) \wedge x \notin C \wedge y \notin C$. By abuse of notation, we let $\mathrm{D}\Gamma_j$ and $\mathrm{D}\Gamma_j^{\times}$ also denote the Cartesian powers of these relations.

If F is a field of meromorphic functions of variables $t_1, \ldots, t_m$ over some complex domain with derivations d/dt_k, then $\mathrm{D}\Gamma_j^{\times}(F)$ is interpreted as the set of all tuples $(x, y) \in F^2$ where $x = x(t_1, \ldots, t_m)$ is some meromorphic function and $y = j(gx)$ for some $g \in \mathrm{GL}_2(\mathbb{C})$.

Pila and Tsimerman proved the following analogue of Ax's theorem for the j-function.

Theorem 3.8 (Ax–Schanuel for j [Pila and Tsimerman 2016, Theorem 1.3]). *Let $(F; +, \cdot, D_1, \ldots, D_m)$ be a differential field with commuting derivations and with field of constants C. Let also $(z_i, j_i) \in \mathrm{D}\Gamma_j^{\times}(F)$, $i = 1, \ldots, n$. If the j_i's are pairwise modularly independent (i.e., no two of them satisfy an equation given by a modular polynomial) then* $\mathrm{td}_C\, C(\bar{z}, \bar{j}) \geq n + \mathrm{rk}(D_k z_i)_{i,k}$.

The proof of Pila and Tsimerman relies on o-minimality and, in particular, the definability of the restriction of the j-function to a fundamental domain in the o-minimal structure $\mathbb{R}_{\mathrm{an,exp}}$. Recently, a differential-algebraic proof of Ax–Schanuel for all Fuchsian automorphic functions (including j) was given in [Blázquez-Sanz et al. 2021].

In [Aslanyan et al. 2021], Aslanyan, Eterović, and Kirby use the Ax–Schanuel theorem for the j-function to establish an existential closedness result for $\mathrm{D}\Gamma_j$. The proof is differential algebraic, and its advantage is that it treats Ax–Schanuel as a black box without looking into it, as opposed to the approach of [Kirby 2009] where the proof of Ax–Schanuel is appealed to. For that reason the proof works both for exp and j, and is expected to work in any reasonable situation where Ax–Schanuel is known.

Theorem 3.9 (functional MEC [Aslanyan et al. 2021, Theorem 1.1]). *Let F be a differential field, and $V \subseteq F^{2n}$ be a Γ_j-broad variety. Then there is a differential field extension $K \supseteq F$ such that $V(K) \cap \mathrm{D}\Gamma_j(K) \neq \varnothing$. In particular, if F is differentially closed then $V(F) \cap \mathrm{D}\Gamma_j(F) \neq \varnothing$.*

Remark 3.10. In the above theorem the variety V need not be free. However, freeness is a necessary condition in MEC; see Example 2.12.

Also, when V is defined over the constants C and is *strongly Γ_j-broad* (i.e., strict inequalities hold in Definition 3.3 (first bullet point)), we have $V(K) \cap \mathrm{D}\Gamma_j^{\times}(K) \neq \varnothing$; see [Aslanyan et al. 2021, Theorem 1.3].

The Ax–Schanuel theorem can also be used to establish a functional variant of modular Zilber–Pink, which was done by Pila and Tsimerman [2016, Theorem 7.1]. They used tools of o-minimality, while [Aslanyan 2022b, Theorem 5.2] gave a differential-algebraic proof based on Kirby's adaptation of Zilber's proof of weak CIT (see [Kirby 2009, Theorem 4.6]).

Definition 3.11. Let $V \subseteq \mathbb{C}^n$ be an algebraic variety. A *j-atypical subvariety* of V is an irreducible component W of some $V \cap T$, where T is a j-special variety, such that $\dim W > \dim V + \dim T - n$. A j-atypical subvariety W of V is said to be *strongly j-atypical* if no coordinate is constant on W.

Theorem 3.12 (functional MZP [Pila and Tsimerman 2016; Aslanyan 2022b]). *Every algebraic variety $V \subseteq \mathbb{C}^n$ contains only finitely many maximal strongly j-atypical subvarieties.*

Like the MZP conjecture, this theorem can also be stated in several equivalent forms, but we do not present them. See [Aslanyan 2022b] for details.

As in the exponential setting, recent years have seen significant progress towards MEC and MZP. For the state-of-the-art on MZP and its generalisations see [Pila 2022] and references therein. For MEC the reader is referred to [Eterović and Herrero 2021; Aslanyan and Kirby 2022; Gallinaro 2021; Eterović 2022; Eterović and Zhao 2021].

4. Incorporating the derivatives of modular functions

In this section we look at the extensions of MSC, MEC, and MZP to the j-function together with its derivatives. Analogues of MSC and MZP in this setting were considered by Pila in some unpublished notes [2013], and we closely follow him in Section 4A and the beginning of Section 4C. MSC with derivatives is in fact a special case of the Grothendieck–André generalised period conjecture. MEC with derivatives was first proposed in [Aslanyan and Kirby 2022]. In addition to that conjecture we also propose a second, more general MEC with derivatives conjecture here.

Recall that j satisfies a third-order differential equation, so it suffices to consider only the first two derivatives. Adding higher derivatives would not change the problems. One normally works in jet spaces when dealing with j together with its derivatives j', j''. However, as usual, instead of the jet space $J_2\,\mathbb{H}^n \times J_2 Y(1)^n$ we work in $\mathbb{H}^n \times \mathbb{C}^{3n}$. We use $(\bar{x}, \bar{y}, \bar{y}', \bar{y}'')$ to denote the coordinates on this space. We denote the vector function

$$(j, j', j''): \mathbb{H}^n \to \mathbb{C}^{3n}, \quad \bar{z} \mapsto (j(\bar{z}), j'(\bar{z}), j''(\bar{z})),$$

by J, and its graph by Γ_J.

Before proceeding we introduce further notation to be used in the rest of this section.

Notation. Let n be a positive integer, $k \leq n$ and $1 \leq i_1 < \cdots < i_k \leq n$.

- $\Pi_{\bar{i}} : \mathbb{C}^{4n} \to \mathbb{C}^{4k}$ is defined by $(\bar{x}, \bar{y}, \bar{y}', \bar{y}'') \mapsto (\mathrm{pr}_{\bar{i}}\,\bar{x}, \mathrm{pr}_{\bar{i}}\,\bar{y}, \mathrm{pr}_{\bar{i}}\,\bar{y}', \mathrm{pr}_{\bar{i}}\,\bar{y}'')$.
- $\pi_{\bar{i}} : \mathbb{C}^{3n} \to \mathbb{C}^{3k}$ is defined by $(\bar{y}, \bar{y}', \bar{y}'') \mapsto (\mathrm{pr}_{\bar{i}}\,\bar{y}, \mathrm{pr}_{\bar{i}}\,\bar{y}', \mathrm{pr}_{\bar{i}}\,\bar{y}'')$.
- We also define the maps

$$\begin{aligned} \pi_{\bar{y}} &: \mathbb{C}^{3n} \to \mathbb{C}^n, && (\bar{y}, \bar{y}', \bar{y}'') \mapsto \bar{y}, \\ \Pi_{\bar{y}} &: \mathbb{C}^{4n} \to \mathbb{C}^n, && (\bar{x}, \bar{y}, \bar{y}', \bar{y}'') \mapsto \bar{y}, \\ \Pi_{\bar{x}} &: \mathbb{C}^{4n} \to \mathbb{C}^n, && (\bar{x}, \bar{y}, \bar{y}', \bar{y}'') \mapsto \bar{x}. \end{aligned}$$

4A. *Modular Schanuel conjecture with derivatives.*

Conjecture 4.1 (modular Schanuel conjecture with derivatives — MSCD). *Given nonquadratic numbers $z_1, \ldots, z_n \in \mathbb{H}$ with distinct $\mathrm{GL}_2^+(\mathbb{Q})$-orbits, we have*

$$\mathrm{td}_{\mathbb{Q}}\,\mathbb{Q}(z_1, \ldots, z_n, J(z_1), \ldots, J(z_n)) \geq 3n.$$

This conjecture is a direct generalisation of MSC, but it does not reflect the transcendence properties of J at special points. So, following [Pila 2013], we formulate a more general conjecture.

Definition 4.2. • An irreducible subvariety $U \subseteq \mathbb{H}^n$ (i.e., an intersection of $\mathbb{H}^n$ with some algebraic variety) is called $\mathrm{GL}_2^+(\mathbb{Q})$-*special* if it is defined by some equations of the form $z_i = g_{i,k} z_k$, $i \neq k$, with $g_{i,k} \in \mathrm{GL}_2^+(\mathbb{Q})$, or of the form $z_i = \tau_i$ where $\tau_i \in \mathbb{H}$ is a quadratic number.

• For a $\mathrm{GL}_2^+(\mathbb{Q})$-special variety U we denote by $\langle U \rangle$ the Zariski closure of the graph of the restriction $J|_U$ (i.e., the set $\{(\bar{z}, J(\bar{z})) : \bar{z} \in U\}$) over $\mathbb{Q}^{\mathrm{alg}}$.

• The $\mathrm{GL}_2^+(\mathbb{Q})$-*special closure* of an irreducible variety $W \subseteq \mathbb{H}^n$ is the smallest $\mathrm{GL}_2^+(\mathbb{Q})$-special variety containing W. It exists because the irreducible components of an intersection of $\mathrm{GL}_2^+(\mathbb{Q})$-special varieties is $\mathrm{GL}_2^+(\mathbb{Q})$-special.

We now explain how $\langle U\rangle$ can be defined algebraically. First let us ignore the case when U has constant coordinates. Assume the first two coordinates of U are related, i.e., $x_2 = gx_1$ for some $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$, and let $\Phi(j(z), j(gz)) = 0$ for some modular polynomial Φ. Differentiating the last equality with respect to z we get

$$\frac{\partial\Phi}{\partial Y_1}(j(z), j(gz)) \cdot j'(z) + \frac{\partial\Phi}{\partial Y_2}(j(z), j(gz)) \cdot j'(gz) \cdot \frac{ad-bc}{(cz+d)^2} = 0. \qquad (\star)$$

Thus, $\langle U\rangle$ satisfies the equation

$$\frac{\partial\Phi}{\partial Y_1}(y_1, y_2) \cdot y_1' + \frac{\partial\Phi}{\partial Y_2}(y_1, y_2) \cdot y_2' \cdot \frac{ad-bc}{(cx_1+d)^2} = 0. \qquad (\dagger)$$

Differentiating again, we get another equation between $(x_1, x_2, y_1, y_2, y_1', y_2', y_1'', y_2'')$, and we have four equations defining the projection of $\langle U\rangle$ to the first two coordinates.

In general, we have a partition of $\{1, \dots, n\}$, where two indices are in the same block of the partition if and only if the corresponding coordinates are related on U. If $i_1 < \cdots < i_k$ form such a block, then $\Pi_{\bar{i}}\langle U\rangle$ is referred to as a *block* of $\langle U\rangle$. Then each block of $\langle U\rangle$ is defined by equations of the form described above and has dimension 4, and $\langle U\rangle$ is the product of its blocks.

When U has a constant coordinate, say x_1 (whose value must be a quadratic irrational), then we also get blocks of dimension 1 or 0 as follows. If

$$x_1 = \tau \notin \mathrm{SL}_2(\mathbb{Z})i \cup \mathrm{SL}_2(\mathbb{Z})\rho, \quad \text{where } \rho = -\frac{1}{2} + \frac{\sqrt{3}}{2}i,$$

then $j(\tau) \in \mathbb{Q}^{\mathrm{alg}}$ and $\mathrm{td}_{\mathbb{Q}}\, \mathbb{Q}(j'(\tau), j''(\tau)) = 1$ (see [Diaz 2000]). If, in addition, some other coordinates, say $x_2, \dots, x_k$, are $\mathrm{GL}_2^+(\mathbb{Q})$-related to x_1 and thus take constant values τ_k (with $\tau_1 := \tau$), then $\mathrm{td}_{\mathbb{Q}}(\bar{\tau}, J(\bar{\tau})) = 1$. Thus, we get a block of dimension 1. The equations defining such a block can be worked out as above.

On the other hand, a constant coordinate in $\mathrm{SL}_2(\mathbb{Z})\rho$ would give rise to a block of dimension 0, for the values of j, j', j'' are zeroes at these points. A constant coordinate in $\mathrm{GL}_2^+(\mathbb{Q})\rho \setminus \mathrm{SL}_2(\mathbb{Z})\rho$ or $\mathrm{GL}_2^+(\mathbb{Q})i$ gives a block of dimension 1.

Now we are ready to state the second (and more general) version of MSCD.

Conjecture 4.3 (modular Schanuel conjecture with derivatives and special points — MSCDS). *Let* $z_1, \dots, z_n \in \mathbb{H}$ *be arbitrary and let* $U \subseteq \mathbb{H}^n$ *be the* $\mathrm{GL}_2^+(\mathbb{Q})$*-special closure of* $(z_1, \dots, z_n)$. *Then*

$$\mathrm{td}_{\mathbb{Q}}\, \mathbb{Q}(z_1, \dots, z_n, J(z_1), \dots, J(z_n)) \geq \dim\langle U\rangle - \dim U.$$

Both MSCD and MSCDS are special cases of the Grothendieck–André generalised period conjecture; see [Aslanyan et al. 2023a, §6.3].

4B. ***Modular existential closedness with derivatives.*** We now introduce the appropriate notions of broadness and freeness which will appear in existential closedness.

Definition 4.4. Let $V \subseteq \mathbb{H}^n \times \mathbb{C}^{3n}$ be an algebraic variety.

- V is *Γ_J-broad* if for any $1 \leq i_1 < \cdots < i_k \leq n$ we have $\dim \Pi_{\bar{\imath}}(V) \geq 3k$.
- V is *modularly free* if no equation of the form $\Phi_N(y_i, y_k) = 0$, or of the form $y_i = c$ with $c \in \mathbb{C}$ a constant, holds on V.
- V is *$\mathrm{GL}_2^+(\mathbb{Q})$-free* if no equation of the form $x_i = gx_k$ with $g \in \mathrm{GL}_2^+(\mathbb{Q})$, or of the form $x_k = c$ with $c \in \mathbb{H}$ a constant, holds on V.
- V is *Γ_J-free* if it is $\mathrm{GL}_2^+(\mathbb{Q})$-free and modularly free.
- V is *Γ_J-froad* if it is Γ_J-free and Γ_J-broad.

Conjecture 4.5 (modular existential closedness with derivatives — MECD). *Let $V \subseteq \mathbb{H}^n \times \mathbb{C}^{3n}$ be a Γ_J-froad variety defined over $\mathbb{C}$. Then $V \cap \Gamma_J \neq \varnothing$.*

This is dual to Conjecture 4.1. It is possible to state a dual to Conjecture 4.3, which would also imply that certain varieties contain J-special points. However, in that case only dimension conditions would not suffice to guarantee existence of J-points, e.g., an arbitrary variety of dimension 1 may not contain such a point; it should be J-special in order to contain J-special points. So we give the following definitions.

Definition 4.6. Let $V \subseteq \mathbb{H}^n \times \mathbb{C}^{3n}$ be an irreducible variety. Let also $U \subseteq \mathbb{H}^n$ be the $\mathrm{GL}_2^+(\mathbb{Q})$-special closure of $\Pi_{\bar{x}}(V)$ and $T \subseteq \mathbb{C}^n$ be the j-special closure[4] of $\Pi_{\bar{y}}(V)$.

- V is said to be *Γ_J^*-free* if $j(U) = T$ and $V \subseteq \langle U \rangle$.
- V is said to be *Γ_J^*-broad* if $\dim \Pi_{\bar{\imath}}(V) \geq \dim \langle \mathrm{pr}_{\bar{\imath}}\, U \rangle - \dim \mathrm{pr}_{\bar{\imath}}\, U$ for any $\bar{\imath}$.
- V is said to be *Γ_J^*-froad* if it is Γ_J^*-free and Γ_J^*-broad.

Remark 4.7. Γ_J^*-freeness means that the $\mathrm{GL}_2^+(\mathbb{Q})$-relations and modular relations holding on V match each other, i.e., are compatible with the functional equations of J (that is, modular correspondences and the relations obtained by differentiating those). This condition holds vacuously for Γ_J-free varieties. For Γ_J-free varieties Γ_J-broadness and Γ_J^*-broadness are equivalent.

Conjecture 4.8 (modular existential closedness with derivatives and special points — MECDS). *Let $V \subseteq \mathbb{H}^n \times \mathbb{C}^{3n}$ be an irreducible Γ_J^*-froad variety. Then $V \cap \Gamma_J \neq \varnothing$.*

[4]The j-special closure of an irreducible set W is defined as the smallest j-special set containing W.

4C. *Modular Zilber–Pink with derivatives.*

Definition 4.9. • For a $\mathrm{GL}_2^+(\mathbb{Q})$-special variety $U \subseteq \mathbb{H}^n$ we denote by $\langle\langle U \rangle\rangle$ the Zariski closure of $J(U)$ over $\mathbb{Q}^{\mathrm{alg}}$.

• A J-*special* subvariety of $\mathbb{C}^{3n}$ is a set of the form $S = \langle\langle U \rangle\rangle$, where U is a $\mathrm{GL}_2^+(\mathbb{Q})$-special subvariety of $\mathbb{H}^n$.

• A J-special variety S is said to be *associated to* a j-special variety T if there is a $\mathrm{GL}_2^+(\mathbb{Q})$-special variety U such that $S = \langle\langle U \rangle\rangle$ and $j(U) = T$.

Remark 4.10. • For a $\mathrm{GL}_2^+(\mathbb{Q})$-special variety $U \subseteq \mathbb{H}^n$ the set $j(U) \subseteq \mathbb{C}^n$ is defined by modular equations and is irreducible (since U is irreducible), and therefore it is j-special. Similarly, $J(U)$ is an irreducible locally analytic set[5] and hence so is its Zariski closure. Thus, J-special varieties are irreducible.

• The j-special varieties are bi-algebraic for the j-function, that is, they are the images under j of algebraic varieties (namely, $\mathrm{GL}_2^+(\mathbb{Q})$-special varieties). That is in contrast to J-special varieties as these are not bi-algebraic for J. Nonetheless, J-special varieties still capture the algebraic properties of the function J.

• The equations defining a J-special variety can be worked out as in Section 4A, since $\langle\langle U \rangle\rangle$ is a projection of $\langle U \rangle$. In particular, a variety $\langle\langle U \rangle\rangle$ is the product of its *blocks* each of which has dimension 0, 1, 3, or 4. Dimensions 0 and 1 correspond to constant coordinates. A block has dimension 3 if all the $\mathrm{GL}_2^+(\mathbb{Q})$-matrices linking its x-coordinates are upper triangular, and dimension 4 otherwise. This is because equation (†) gives an algebraic relation between y_1, y_2, y_1', y_2' provided that $c = 0$, i.e., the matrix linking x_1 and x_2 is upper triangular. Then we also have another such equation linking $y_1, y_2, y_1', y_2', y_1'', y_2''$ obtained by differentiating ($\star$). When $c \neq 0$, both of these equations depend on x_1, so together they yield a single algebraic relation between $y_1, y_2, y_1', y_2', y_1'', y_2''$.

Definition 4.11. For a variety $V \subseteq \mathbb{C}^{3n}$ we let the J-*atypical set* of V, denoted $\mathrm{Atyp}_J(V)$, be the union of all atypical components of intersections $V \cap T$ in $\mathbb{C}^{3n}$, where $T \subseteq \mathbb{C}^{3n}$ is a J-special variety.

Conjecture 4.12 (modular Zilber–Pink with derivatives — MZPD [Pila 2013]). *For every algebraic variety $V \subseteq \mathbb{C}^{3n}$ there is a finite collection Σ of proper $\mathrm{GL}_2^+(\mathbb{Q})$-special subvarieties of $\mathbb{H}^n$ such that*

$$\mathrm{Atyp}_J(V) \cap J(\mathbb{H}^n) \subseteq \bigcup_{\substack{U \in \Sigma \\ \bar{\gamma} \in \mathrm{SL}_2(\mathbb{Z})^n}} \langle\langle \bar{\gamma} U \rangle\rangle.$$

[5] Strictly speaking, $J(U)$ may not be complex analytic as it is the image of an analytic set under an analytic function, but it is locally analytic. It is irreducible in the sense that if $J(U)$ is contained in a countable union of analytic sets then it must be contained in one of them.

Remark 4.13. • One could propose a stronger conjecture stating that $\mathrm{Atyp}_J(V)$ is covered by J-special varieties corresponding to $\mathrm{SL}_2(\mathbb{Z})$-translates of finitely many $\mathrm{GL}_2^+(\mathbb{Q})$-special varieties $U \in \Sigma$. However, an intersection of V with a J-special variety may have a component which does not intersect the image of J, or that intersection is small. So while this stronger statement seems sensible (meaning there does not seem to be a trivial counterexample), it is less natural and less about the function J than Conjecture 4.12. Zilber's original motivation for CIT came from the idea of deducing a uniform version of Schanuel from itself. Similarly, [Pila 2013] proposes MZPD as the difference between MSCD and its uniform version. Since MSCD is about the function J, Pila only needed to deal with the part of $\mathrm{Atyp}_J(V)$ that consists of points from the image of J. Furthermore, Conjecture 4.12 is supported by the theorems presented in Section 4D, while we do not have any evidence towards the said stronger statement, so we do not propose such a conjecture.

• Given a J-special variety $S \subseteq \mathbb{C}^{3n}$ with an atypical intersection $V \cap S$, the intersection $\pi_{\bar{y}}(V) \cap \pi_{\bar{y}}(S)$ may or may not be atypical. The novelty of MZPD is when this intersection is typical as the atypical ones are accounted for MZP.

• In MZPD we may need infinitely many J-special varieties to cover the set $\mathrm{Atyp}_J(V) \cap J(\mathbb{H}^n)$ but the conjecture states that it is "generated" by finitely many $\mathrm{GL}_2^+(\mathbb{Q})$-special varieties. See the example below.

Example 4.14. Consider the variety $V \subseteq \mathbb{C}^9$ defined by $\Phi_2(y_1, y_2)+\Phi_3(y_2, y_3)=0$. Let $T \subseteq \mathbb{C}^3$ be a j-special variety defined by $\Phi_2(y_1, y_2) = \Phi_3(y_2, y_3) = 0$, and let $U \subseteq \mathbb{H}^3$ be $\mathrm{GL}_2^+(\mathbb{Q})$-special such that $j(U) = T$. Then for every $\bar{\gamma} \in (\mathrm{SL}_2(\mathbb{Z}))^3$ we have $\langle\langle \bar{\gamma} U \rangle\rangle \subseteq V$, and these are maximal J-special (hence atypical) in V. Thus, the single j-special variety T "generates" an infinite set of maximal J-atypical subvarieties of V.

MZPD has an analytic component: the intersection of $\mathrm{Atyp}_J(V)$ with the image of J. We now propose an "algebraic" MZPD conjecture which we believe will be more amenable to (differential) algebraic and geometric techniques (below we provide evidence in support of this). The idea is to replace the set of points from the image of J in an atypical subvariety of V by its Zariski closure. Then we need to understand which algebraic varieties can contain a Zariski dense set of such points, and hence this is a variant of the existential closedness problem for J. So we define an appropriate notion of froadness which serves that purpose.

Definition 4.15. An irreducible variety $W \subseteq \mathbb{C}^{3n}$ is called $\mathrm{Im}(J)$-*froad* (resp. $\mathrm{Im}(J)^*$-*froad*)[6] if it is the projection of a Γ_J-froad (resp. Γ_J^*-froad) variety $V \subseteq \mathbb{H}^n \times \mathbb{C}^{3n}$ to the coordinates $(\bar{y}, \bar{y}', \bar{y}'')$.

[6] Here Im stands for the image of a function.

The following statement gives an explicit definition of these notions. Its proof is fairly straightforward from the definitions and is left to the reader.

Proposition 4.16. *Let $W \subseteq \mathbb{C}^{3n}$ be an irreducible variety, and let $T \subseteq \mathbb{C}^n$ be the j-special closure of $\pi_{\bar{y}}(W)$. Then W is* $\mathrm{Im}(J)^*$*-froad if and only if there is a* $\mathrm{GL}_2^+(\mathbb{Q})$*-special variety $U \subseteq \mathbb{H}^n$ such that*

- $j(U) = T$,
- $W \subseteq \langle\langle U \rangle\rangle$,
- *for any $\bar{\imath}$ we have* $\dim \pi_{\bar{\imath}}(W) \geq \dim \langle\langle \mathrm{pr}_{\bar{\imath}}(U) \rangle\rangle - \dim \mathrm{pr}_{\bar{\imath}}(U)$.

Furthermore, W is $\mathrm{Im}(J)$*-froad if and only if $U = \mathbb{H}^n$, $T = \mathbb{C}^n$, and for any $\bar{\imath}$ of length k we have* $\dim \pi_{\bar{\imath}}(W) \geq 2k$.

Definition 4.17. For a variety $V \subseteq \mathbb{C}^{3n}$ we let the *froadly J-atypical set* of V, denoted $\mathrm{FAtyp}_J(V)$, be the union of all $\mathrm{Im}(J)^*$-froad and atypical components of intersections $V \cap T$ in $\mathbb{C}^{3n}$, where $T \subseteq \mathbb{C}^{3n}$ is a J-special variety.

Conjecture 4.18 (modular Zilber–Pink with derivatives for froad varieties — MZPDF). *For every algebraic variety $V \subseteq \mathbb{C}^{3n}$ there is a finite collection Σ of proper* $\mathrm{GL}_2^+(\mathbb{Q})$*-special subvarieties of $\mathbb{H}^n$ such that*

$$\mathrm{FAtyp}_J(V) \subseteq \bigcup_{\substack{U \in \Sigma \\ \bar{\gamma} \in \mathrm{SL}_2(\mathbb{Z})^n}} \langle\langle \bar{\gamma} U \rangle\rangle.$$

Now we aim to understand the relation between Conjectures 4.12 and 4.18. We can show they are equivalent assuming some weakened versions of MSCD and MECD referring only to the image of J. We call these conjectures MSCDI and MECDI, where “I” stands for “Image”.

Conjecture 4.19 (MSCDI). *Let $z_1, \dots, z_n \in \mathbb{H}$ be arbitrary and let $U \subseteq \mathbb{H}^n$ be the* $\mathrm{GL}_2^+(\mathbb{Q})$*-special closure of $(z_1, \dots, z_n)$. Then*

$$\mathrm{td}_{\mathbb{Q}}\, \mathbb{Q}(J(z_1), \dots, J(z_n)) \geq \dim\langle\langle U \rangle\rangle - \dim U.$$

Conjecture 4.20 (MECDI). *Let $V \subseteq \mathbb{C}^{3n}$ be an irreducible* $\mathrm{Im}(J)^*$*-froad variety. Then $V \cap \mathrm{Im}(J) \neq \varnothing$.*

Proposition 4.21. (i) *Assume MECDI. Then Conjecture 4.12 (MZPD) implies Conjecture 4.18 (MZPDF).*

(ii) *Assume MSCDI. Then Conjecture 4.18 (MZPDF) implies Conjecture 4.12 (MZPD).*

Proof. (i) Let W be an $\mathrm{Im}(J)^*$-froad atypical subvariety of $V \subseteq \mathbb{C}^{3n}$. Then by MECDI and the Rabinowitsch trick (see [Aslanyan 2022a, Proposition 4.34]), $W \cap \mathrm{Im}(J)$ is Zariski dense in W. By MZPD (Conjecture 4.12), $\pi_{\bar{y}}(W \cap \mathrm{Im}(J))$

is contained in a union of finitely many j-special varieties depending only on V. Hence, $\pi_{\bar{y}}(W) = \overline{\pi_{\bar{y}}(W \cap \mathrm{Im}(J))}^{\mathrm{Zcl}}$ is also contained in that union. Since W is irreducible, $\pi_{\bar{y}}(W)$ is contained in one such j-special variety T, and since W is $\mathrm{Im}(J)^*$-froad, it is contained in a J-special variety associated to T.

(ii) (cf. [Aslanyan 2022b, Proposition 9.10]) Now assume MSCDI and MZPDF. Also assume first that V is defined over $\mathbb{Q}^{\mathrm{alg}}$. Let $\bar{w} := (j(\bar{z}), j'(\bar{z}), j''(\bar{z})) \in \mathrm{Atyp}_J(V)$ belong to an atypical component of an intersection $V \cap T$, where T is J-special. If $T' \subseteq T$ is the J-special closure of $\bar{w}$ (that is, $T' = \langle\langle U \rangle\rangle$, where U is the $\mathrm{GL}_2^+(\mathbb{Q})$-special closure of $\bar{z}$), then by [Aslanyan 2022b, Lemma 9.9], $\bar{w}$ belongs to an atypical component W of the intersection $V \cap T'$. MSCDI implies that W is $\mathrm{Im}(J)^*$-froad. Hence, by MZPDF W is contained in a J-special variety S associated to one of the finitely many j-special varieties depending only on V. Then $\bar{w}$ also belongs to S.

When V is defined over arbitrary parameters, rather than $\mathbb{Q}^{\mathrm{alg}}$, the same proof goes through provided that we can extend MSCDI and get a lower bound on the transcendence degree of a J-point over finitely generated fields. This has been done in [Aslanyan et al. 2023a, §5] for MSCD (see also [Eterović 2022, §4.2]), and MSCDI can be treated similarly. □

MSCDI, like full MSCD, seems to be out of reach. Hence the second part of the above proposition is not very helpful. On the other hand, MECDI is within reach, albeit still open. Therefore, the first part of the proposition is more meaningful and tells us that MZPDF (Conjecture 4.18) is probably more tractable than MZPD (Conjecture 4.12). It is unlikely that the second implication in Proposition 4.21 can be proven without assuming MSCDI.

4D. ***Functional/differential variants.*** The functional variants of all the above conjectures were established in the last decade. We present them below.

Definition 4.22. Let $(F; +, \cdot, D_1, \dots, D_m)$ be a differential field with constant field $C = \bigcap_{k=1}^m \ker D_k$. Let also Ψ be the rational function appearing in the differential equation of the j-function (see Section 3C).

- We define a 4-ary relation $\mathrm{D}\Gamma_J(x, y, y', y'')$ by

$$\exists y''' \Big[\Psi(y, y', y'', y''') = 0 \wedge \bigwedge_{k=1}^m D_k y = y' D_k x \wedge D_k y' = y'' D_k x \wedge D_k y'' = y''' D_k x \Big].$$

- The relation $\mathrm{D}\Gamma_J^{\times}(x, y, y', y'')$ is defined by the formula

$$\mathrm{D}\Gamma_J(x, y, y', y'') \wedge x \notin C \wedge y \notin C \wedge y' \notin C \wedge y'' \notin C.$$

- The relations $\mathrm{DIm}(J)$ and $\mathrm{DIm}(J)^{\times}$ are defined as $\exists x\, \mathrm{D}\Gamma_J(x, y, y', y'')$ and $\exists x\, \mathrm{D}\Gamma_J^{\times}(x, y, y', y'')$, respectively.
- By abuse of notation, we use the above expressions ($\mathrm{D}\Gamma_J$, $\mathrm{D}\Gamma_J^{\times}$, etc.) to denote their Cartesian powers too.

If F is a field of meromorphic functions of variables $t_1, \dots, t_m$ over some complex domain with derivations d/dt_k, then $\mathrm{D}\Gamma_J^{\times}(F)$ is interpreted as the set of all tuples $(x, y, y', y'') \in F^4$ where $x = x(t_1, \dots, t_m)$ is some meromorphic function, $y = j(gx)$ for some $g \in \mathrm{GL}_2(\mathbb{C})$, and $y' = dj(gx)/dx$, $y'' = d^2 j(gx)/dx^2$.

The Ax–Schanuel theorem for J is due to Pila and Tsimerman. Again, their proof is based on o-minimality, and Blázquez-Sanz, Casale, Freitag, and Nagloo give a differential-algebraic/model-theoretic proof in [Blázquez-Sanz et al. 2021].

Theorem 4.23 (Ax–Schanuel for J [Pila and Tsimerman 2016, Theorem 1.3]). *Let $(F; +, \cdot, D_1, \dots, D_m)$ be a differential field with commuting derivations and with field of constants C. Let also $(z_i, j_i, j_i', j_i'') \in \mathrm{D}\Gamma_J^{\times}(F)$, $i = 1, \dots, n$. If the j_i are pairwise modularly independent then* $\mathrm{td}_C\, C(\bar{z}, \bar{j}, \bar{j}', \bar{j}'') \geq 3n + \mathrm{rk}(D_k z_i)_{i,k}$.

As in the previous section, Ax–Schanuel can be used to prove a differential analogue of MECD.

Theorem 4.24 (differential MECD [Aslanyan et al. 2021, Theorem 1.2]). *Let F be a differential field, and $V \subseteq F^{4n}$ be a Γ_J-broad variety. Then there is a differential field extension $K \supseteq F$ such that $V(K) \cap \mathrm{D}\Gamma_J(K) \neq \varnothing$. In particular, if F is differentially closed then $V(F) \cap \mathrm{D}\Gamma_J(F) \neq \varnothing$.*

Remark 4.25. In this theorem, when V is defined over the constants C and is *strongly Γ_J-broad* (i.e., strict inequalities hold in Definition 4.4 (first bullet point)), we have $V(K) \cap \mathrm{D}\Gamma_j^{\times}(K) \neq \varnothing$; see [Aslanyan et al. 2021, Theorem 1.3].

At the end we state several analogues of MZPD and MZPDF.

Definition 4.26. For a J-special variety $T \subseteq \mathbb{C}^{3n}$ and an algebraic variety $V \subseteq \mathbb{C}^{3n}$ an atypical component W of an intersection $V \cap T$ in $\mathbb{C}^{3n}$ is a *strongly J-atypical* subvariety of V if for every irreducible analytic component W_0 of $W \cap J(\mathbb{H}^n)$, no coordinate is constant on $\pi_{\bar{y}}(W_0)$. The *strongly J-atypical set* of V, denoted $\mathrm{SAtyp}_J(V)$, is the union of all strongly J-atypical subvarieties of V.

The following is a weak version of MZPD, the proof of which is based on complex geometric tools. It generalises functional MZP (Theorem 3.12), and hence it gives a third proof of the latter.

Theorem 4.27 (weak MZPD [Aslanyan 2022b, Theorem 7.9]). *For every algebraic variety $V \subseteq \mathbb{C}^{3n}$ there is a finite collection Σ of proper $\mathrm{GL}_2^+(\mathbb{Q})$-special subvarieties of $\mathbb{H}^n$ such that*

$$\mathrm{SAtyp}_J(V) \cap J(\mathbb{H}^n) \subseteq \bigcup_{\substack{U \in \Sigma \\ \bar{\gamma} \in \mathrm{SL}_2(\mathbb{Z})^n}} \langle\langle \bar{\gamma} U \rangle\rangle.$$

In order to present differential analogues of MZPD(F), we need to introduce several definitions and pieces of notation.

Definition 4.28 [Aslanyan 2022b, §6]. Let C be an algebraically closed field. Define D as the zero derivation on C and extend $(C; +, \cdot, D)$ to a differentially closed field $(K; +, \cdot, D)$.

- Let $T \subseteq C^n$ be a j-special variety and $U \subseteq C^n$ be a $\mathrm{GL}_2(C)$-special variety associated to T, that is, U is defined by $\mathrm{GL}_2(C)$-equations and for any i, k the pair of coordinates x_i, x_k are related on U if and only if y_i, y_k are modularly related on T. Denote by $\langle\langle U, T\rangle\rangle$ the Zariski closure over C of the projection of the set
$$\mathrm{D}\Gamma_J^{\times}(K) \cap (U(K) \times T(K) \times K^2)$$
to the coordinates $(\bar{y}, \bar{y}', \bar{y}'')$.
- A D_J-*special* variety is a variety $S := \langle\langle U, T\rangle\rangle$ for some T and U as above.
- $S \sim T$ means that $S := \langle\langle U, T\rangle\rangle$ for some U associated to T. For a set Σ of j-special varieties $S \sim \Sigma$ means that $S \sim T$ for some $T \in \Sigma$.

Definition 4.29. Let $V \subseteq C^{3n}$ be a variety. The D_J-*atypical set* of V, denoted $\mathrm{Atyp}_{\mathrm{D}_J}(V)$, is the union of all D_J-atypical subvarieties of V, that is, atypical components of intersections $V \cap T$, where $T \subseteq C^{3n}$ is D_J-special. The set $\mathrm{SFAtyp}_{\mathrm{D}_J}(V)$ denotes the union of all D_J-atypical subvarieties of V which are strongly $\mathrm{Im}(J)^*$-froad.[7]

Theorem 4.30 (functional MZPD — FMZPD [Aslanyan 2022b, Theorem 8.2]). *Let $(K; +, \cdot, D)$ be a differential field with an algebraically closed field of constants C. Given an algebraic variety $V \subseteq C^{3n}$, there is a finite collection Σ of proper j-special subvarieties of C^n such that*
$$\mathrm{Atyp}_{\mathrm{D}_J}(V)(K) \cap \mathrm{DIm}_J^{\times}(K) \subseteq \bigcup_{S \sim \Sigma} S.$$

Theorem 4.31 (functional MZPDF — FMZPDF [Aslanyan 2022b, Theorem 9.8]). *Let C be an algebraically closed field of characteristic* 0. *Given an algebraic variety $V \subseteq C^{3n}$, there is a finite collection Σ of proper j-special subvarieties of C^n such that*
$$\mathrm{SFAtyp}_{\mathrm{D}_J}(V)(C) \subseteq \bigcup_{S \sim \Sigma} S.$$

These theorems are analogues of MZPD and MZPDF, respectively, and so they support those conjectures. In [Aslanyan 2022b] we give a complex geometric proof of FMZPD (the transition from complex geometry to differential algebra is via

[7]A variety $W \subseteq \mathbb{C}^{3n}$ is *strongly* $\mathrm{Im}(J)^*$-*froad* if there is a $\mathrm{GL}_2^+(\mathbb{Q})$-special variety $U \subseteq \mathbb{H}^n$ such that $j(U) = T$, $W \subseteq \langle\langle U\rangle\rangle$, and for any $\bar{\imath}$ we have $\dim \pi_{\bar{\imath}}(W) > \dim \langle\langle \mathrm{pr}_{\bar{\imath}}(U)\rangle\rangle - \dim \mathrm{pr}_{\bar{\imath}}(U)$.

Seidenberg's embedding theorem) and a differential-algebraic proof of FMZPDF. The core of both proofs is the Ax–Schanuel theorem for J. The proof of FMZPDF also uses the differential version of MECDI, which is a special case of Theorem 4.24. As above, FMZPD and FMZPDF can be deduced from one another using differential MECDI, so that gives two proofs for each of the above theorems, one differential algebraic and one complex geometric.

For further results on MECD and MZPD see [Aslanyan and Kirby 2022; Aslanyan et al. 2023a; Eterović 2022] and [Aslanyan 2022b], respectively. Spence [2019] has proven some results towards the modular André–Oort with derivatives conjecture, which is a special case of MZPD.

Remark 4.32. Section 4 turned out to be somewhat technical with some hard-to-remember concepts and notation. Unfortunately, that seems necessary for precision and rigour. A reader who is not familiar with the general topics discussed here may be lost in the various versions of the conjectures and theorems. Therefore, we would like to reiterate the main high-level idea of this section: incorporating the derivatives into the modular versions of the Schanuel, existential closedness, and Zilber–Pink conjectures gives a deeper insight into these problems and reveals some hidden (possibly surprising) connections between them. Exploring these conjectures in this more general setting would allow us to better understand the full model-theoretic picture.

Acknowledgement

I would like to thank the anonymous referees for their useful comments.

References

[André 2004] Y. André, *Une introduction aux motifs* (*motifs purs*, *motifs mixtes*, *périodes*), Panoramas et Synthèses **17**, Soc. Math. France, Paris, 2004. MR Zbl

[Aslanyan 2022a] V. Aslanyan, "Adequate predimension inequalities in differential fields", *Ann. Pure Appl. Logic* **173**:1 (2022), art. id. 103030. MR Zbl

[Aslanyan 2022b] V. Aslanyan, "Weak modular Zilber–Pink with derivatives", *Math. Ann.* **383**:1-2 (2022), 433–474. MR Zbl

[Aslanyan and Kirby 2022] V. Aslanyan and J. Kirby, "Blurrings of the j-function", *Q. J. Math.* **73**:2 (2022), 461–475. MR Zbl

[Aslanyan et al. 2021] V. Aslanyan, S. Eterović, and J. Kirby, "Differential existential closedness for the j-function", *Proc. Amer. Math. Soc.* **149**:4 (2021), 1417–1429. MR Zbl

[Aslanyan et al. 2023a] V. Aslanyan, S. Eterović, and J. Kirby, "A closure operator respecting the modular j-function", *Israel J. Math.* **253**:1 (2023), 321–357. MR Zbl

[Aslanyan et al. 2023b] V. Aslanyan, J. Kirby, and V. Mantova, "A geometric approach to some systems of exponential equations", *Int. Math. Res. Not.* **2023**:5 (2023), 4046–4081. MR Zbl

[Ax 1971] J. Ax, "On Schanuel's conjectures", *Ann. of Math.* (2) **93** (1971), 252–268. MR Zbl

[Bays and Kirby 2018] M. Bays and J. Kirby, "Pseudo-exponential maps, variants, and quasiminimality", *Algebra Number Theory* **12**:3 (2018), 493–549. MR Zbl

[Bertolin 2002] C. Bertolin, "Périodes de 1-motifs et transcendance", *J. Number Theory* **97**:2 (2002), 204–221. MR Zbl

[Blázquez-Sanz et al. 2021] D. Blázquez-Sanz, G. Casale, J. Freitag, and J. Nagloo, "A differential approach to Ax–Schanuel, I", 2021. arXiv 2102.03384

[Bombieri et al. 2007] E. Bombieri, D. Masser, and U. Zannier, "Anomalous subvarieties—structure theorems and applications", *Int. Math. Res. Not.* **2007**:19 (2007), art. id. rnm057. MR Zbl

[Brownawell and Masser 2017] W. D. Brownawell and D. W. Masser, "Zero estimates with moving targets", *J. Lond. Math. Soc.* (2) **95**:2 (2017), 441–454. MR Zbl

[D'Aquino et al. 2018] P. D'Aquino, A. Fornasiero, and G. Terzo, "Generic solutions of equations with iterated exponentials", *Trans. Amer. Math. Soc.* **370**:2 (2018), 1393–1407. MR Zbl

[Diaz 2000] G. Diaz, "Transcendance et indépendance algébrique: liens entre les points de vue elliptique et modulaire", *Ramanujan J.* **4**:2 (2000), 157–199. MR Zbl

[Eterović 2022] S. Eterović, "Generic solutions of equations involving the modular j-function", preprint, 2022. arXiv 2209.12192

[Eterović and Herrero 2021] S. Eterović and S. Herrero, "Solutions of equations involving the modular j function", *Trans. Amer. Math. Soc.* **374**:6 (2021), 3971–3998. MR Zbl

[Eterović and Zhao 2021] S. Eterović and R. Zhao, "Algebraic varieties and automorphic functions", preprint, 2021. arXiv 2107.10392

[Freitag and Scanlon 2018] J. Freitag and T. Scanlon, "Strong minimality and the j-function", *J. Eur. Math. Soc.* (*JEMS*) **20**:1 (2018), 119–136. MR Zbl

[Gallinaro 2021] F. Gallinaro, "Solving systems of equations of raising-to-powers type", preprint, 2021. arXiv 2103.15675

[Gallinaro 2023] F. P. Gallinaro, "Exponential sums equations and tropical geometry", *Selecta Math.* (*N.S.*) **29**:4 (2023), art. id. 49. MR Zbl

[Kirby 2009] J. Kirby, "The theory of the exponential differential equations of semiabelian varieties", *Selecta Math.* (*N.S.*) **15**:3 (2009), 445–486. MR Zbl

[Kirby and Zilber 2014] J. Kirby and B. Zilber, "Exponentially closed fields and the conjecture on intersections with tori", *Ann. Pure Appl. Logic* **165**:11 (2014), 1680–1706. MR Zbl

[Lang 1966] S. Lang, *Introduction to transcendental numbers*, Addison-Wesley, Reading, MA, 1966. MR Zbl

[Mahler 1969] K. Mahler, "On algebraic differential equations satisfied by automorphic functions", *J. Austral. Math. Soc.* **10** (1969), 445–450. MR Zbl

[Pila 2013] J. Pila, "Modular Zilber–Pink with derivatives", preprint, 2013.

[Pila 2022] J. Pila, *Point-counting and the Zilber–Pink conjecture*, Cambridge Tracts Math. **228**, Cambridge Univ. Press, 2022. MR Zbl

[Pila and Tsimerman 2016] J. Pila and J. Tsimerman, "Ax–Schanuel for the j-function", *Duke Math. J.* **165**:13 (2016), 2587–2605. MR Zbl

[Pink 2005a] R. Pink, "A combination of the conjectures of Mordell–Lang and André–Oort", pp. 251–282 in *Geometric methods in algebra and number theory*, edited by F. Bogomolov and Y. Tschinkel, Progr. Math. **235**, Birkhäuser, Boston, 2005. MR Zbl

[Pink 2005b] R. Pink, "A common generalization of the conjectures of André–Oort, Manin–Mumford and Mordell–Lang", preprint, 2005, https://people.math.ethz.ch/~pink/ftp/AOMMML.pdf.

[Spence 2019] H. Spence, "A modular André–Oort statement with derivatives", *Proc. Edinb. Math. Soc.* (2) **62**:2 (2019), 323–365. MR Zbl

[Tsimerman 2015] J. Tsimerman, "Ax–Schanuel and o-minimality", pp. 216–221 in *O-minimality and diophantine geometry*, edited by G. O. Jones and A. J. Wilkie, London Math. Soc. Lecture Note Ser. **421**, Cambridge Univ. Press, 2015. MR Zbl

[Zilber 2002] B. Zilber, "Exponential sums equations and the Schanuel conjecture", *J. London Math. Soc.* (2) **65**:1 (2002), 27–44. MR Zbl

[Zilber 2005] B. Zilber, "Pseudo-exponentiation on algebraically closed fields of characteristic zero", *Ann. Pure Appl. Logic* **132**:1 (2005), 67–95. MR Zbl

[Zilber 2015] B. Zilber, "The theory of exponential sums", preprint, 2015. arXiv 1501.03297

Received 9 Dec 2022. Revised 26 Sep 2023.

VAHAGN ASLANYAN:

vahagn.aslanyan@manchester.ac.uk
Department of Mathematics, University of Manchester, Manchester, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.625

Zilber–Pink for raising to the power i

Jonathan Pila

To Boris Zilber on the occasion of his 75th birthday.

We consider the multivalued raising-to-the-power-i function through the Schanuel–Ax–Zilber lens. We formulate and prove an analogue of the Zilber–Pink conjecture.

1. Introduction

The purpose of this paper is to consider the (multivalued) function $w = z^i$ through the Schanuel–Ax–Zilber lens [Ax 1971; Zilber 2002], and in particular to formulate and prove an analogue of the Zilber–Pink conjecture [Zilber 2002; Bombieri et al. 2007; Pink 2005]. We follow the path taken by Zilber leading to his formulation of the Zilber–Pink conjecture for semiabelian varieties: beginning with an analogue of Schanuel's Conjecture 3.1, we consider a "uniform" version (Conjecture 4.1), and formulate a Zilber–Pink-type statement (Conjecture 4.2) connecting the classical Schanuel conjecture (SC) with the uniform version for z^i. Our Schanuel variant is equivalent to a formulation of Zilber [2003a].

We then prove the Zilber–Pink-type statement, in the more general form in Theorem 1.3 below. The connection with SC is explicated in Sections 2 and 3. Theorem 1.3 is a somewhat exotic variant of the Zilber–Pink conjecture for even powers of $\mathbb{G}_{\mathrm{m}} = \mathbb{G}_{\mathrm{m}}(\mathbb{C}) = \mathbb{C}^\times$, in which key difficulties disappear thanks to the Gelfond–Schneider theorem.

We take $w = z^i$ to be the predicate $\Gamma \subset \mathbb{G}_{\mathrm{m}}^2$ defined by

$$(z, w) \in \Gamma \iff \exists u \in \mathbb{C}\, [\exp(u) = z \wedge \exp(iu) = w].$$

We let $\Gamma_n = \Gamma^n$ denote the cartesian power of this predicate on $\mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$.

To formulate our theorem, we recall that the Zilber–Pink conjecture (ZP) for subvarieties of $\mathbb{G}_{\mathrm{m}}^n$ can be framed in terms of *optimal subvarieties* for $V \subseteq \mathbb{G}_{\mathrm{m}}^n$; see [Habegger and Pila 2016]. These are subvarieties $W \subseteq V$ which cannot be enlarged inside V without increasing their *defect* (which is the difference between their dimension and the dimension of the smallest *torsion coset* of $\mathbb{G}_{\mathrm{m}}^n$ which contains them). ZP is equivalent to the statement that a subvariety $V \subset \mathbb{G}_{\mathrm{m}}^n$ contains only

MSC2020: 03C65, 11J91, 14G99.
Keywords: Zilber–Pink conjecture, Schanuel's conjecture.

finitely many such optimal subvarieties. Here torsion cosets (which are cosets of subtori by torsion points) are the "special subvarieties" of $\mathbb{G}_{\mathrm{m}}^n$.

In treating $w = z^i$, the appropriate "special subvarieties" are subtori of $\mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ of a special form.

Definition 1.1. A *plu-torus* $T \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ is a subtorus whose lattice of exponent vectors $\Lambda(T) \subset \mathbb{Z}^n \times \mathbb{Z}^n$ is closed under the operation $(q, r) \mapsto (-r, q)$.

If $A \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ is a subvariety which meets Γ_n then there is a smallest plu-torus containing A (see Section 2), denoted $((A))$. We define the *plu-defect* of A to be

$$\delta(A) = \dim ((A)) - \dim A.$$

Definition 1.2. Let $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$. A subvariety $A \subset V$ is called *plu-optimal* for V if $A \cap \Gamma_n \neq \varnothing$, and if $A \subset B \subset V$ and $\delta(B) \leq \delta(A)$ imply $B = A$.

Theorem 1.3. *Let $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$. Then V contains only finitely many plu-optimal subvarieties.*

To motivate the above definitions, we consider implications of SC for $w = z^i$. Zilber [2003a; 2015] studied Schanuel-type conjectures for raising to powers in algebraically closed fields, and model-theoretic properties of the fields satisfying them. A version of the present result that SC implies a certain uniformity in the corresponding conjecture for raising to the power i is obtained there. Zilber uses a two-sorted setup, and the conjectures are framed using the corresponding logarithms, whereas our statement involves only the relation Γ. However, the two structures are biinterpretable, and the raising-to-the power-i Schanuel conjectures are equivalent (I thank J. Kirby for explaining these points to me). Structures with a predimension with similar shape to that considered here are considered in [Caycedo and Zilber 2014; Zilber 2003b]; see also related structures in the context of "pseudoexponentiation" discussed in [Bays and Kirby 2018].

We first observe that the pair $(z, w) \in \Gamma$ "knows" which branch of log connects them: if $u = \log z$ has the required property, any other u' would need to satisfy

$$u' - u \in 2\pi i\mathbb{Z}, \qquad iu' - iu \in 2\pi i\mathbb{Z}$$

and the intersection of $2\pi i\mathbb{Z}$ and $2\pi\mathbb{Z}$ consists of $\{0\}$ only.

Applying SC to

$$u_1, \dots, u_n, iu_1, \dots, iu_n, x_1, \dots, x_n, y_1, \dots, y_n,$$

where $\exp(u_i) = x_i$, $\exp(iu_i) = y_i$, and eliminating the u_i, iu_i, gives the following statement, in which $\mathrm{t.d.}(A)$ denotes $\mathrm{tr.deg.}\, \mathbb{Q}(A)/\mathbb{Q}$:

Let $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{C}^\times$ with $(x_j, y_j) \in \Gamma$, $j = 1, \dots, n$. Then

$$\mathrm{t.d.}(x_1, \dots, x_n, y_1, \dots, y_n) \geq n$$

unless $x_1, \dots, x_n, y_1, \dots, y_n$ are multiplicatively dependent.

Indeed, SC asserts that $\text{t.d.}(u_j, iu_j, x_j, y_j : 1 \leq j \leq n) \geq 2n$ unless u_j, iu_j are linearly dependent over $\mathbb{Q}$. Now $\text{t.d.}(u_j, iu_j : 1 \leq j \leq n) \leq n$, while if the u_j, iu_j are linearly dependent over $\mathbb{Q}$ then the x_j, y_j are multiplicatively dependent: if, say, $\sum_j q_j u_j + i \sum_j r_j u_j = 0$, where $q_j, r_j \in \mathbb{Z}$, not all zero, then we get

$$\prod_j x_j^{q_j} \prod_j y_j^{r_j} = 1.$$

However, multiplicative dependence of x_j, y_j might hold even when $\mathbb{Q}$-linear dependence of u_j, iu_j does not, so that the above statement seems to lose some information. For example, if $u_1 = \log 2$, $u_2 = 2\pi i$, giving

$$x_1 = 2, \quad x_2 = 1, \quad y_1 = 2^i, \quad y_2 = e^{-2\pi}$$

then the above conjecture does not predict $\text{t.d.}(x_1, x_2, y_1, y_2) \geq 2$, but SC does, as does Conjecture 3.1 (or see the provisional version below).

Suppose $\sum_j q_j u_j + i \sum_j r_j u_j = 0$ as above. Then, upon multiplying by i, we find that $-\sum_j r_j u_j + i \sum_j q_j u_j = 0$ and we get a second multiplicative relation

$$\prod_j x_j^{-r_j} \prod_j y_j^{q_j} = 1.$$

The claim is that a pair of such multiplicative relations (which is easily seen to never be dependent) forces the underlying u_j, iu_j to be linearly dependent over $\mathbb{Q}$. Indeed, from the first, we find that

$$\delta = \sum_j q_j u_j + i \sum_j r_j u_j \in 2\pi i \mathbb{Z}$$

but then the second relation implies that

$$i\delta = -\sum_j r_j u_j + i \sum_j q_j u_j \in 2\pi i \mathbb{Z},$$

and so $\delta = 0$.

Thus our "exceptional" point $(x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ lies in a subtorus of codimension at least 2 and of rather specific form: a plu-torus. This leads us to the following provisional formulation of Schanuel's conjecture for z^i:

Let $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{C}^\times$ with $(x_j, y_j) \in \Gamma$, $j = 1, \ldots, n$. Then

$$\text{t.d.}(x_1, \ldots, x_n, y_1, \ldots, y_n) \geq n$$

unless there exist integers q_j, r_j, not all zero, such that

$$\prod_j x_j^{q_j} \prod_j y_j^{r_j} = 1 = \prod_j x_j^{-r_j} \prod_j y_j^{q_j}.$$

Thus, SC for z^i leads naturally to the consideration of plu-tori. The corresponding Zilber–Pink analogue arises from considering a uniform version of SC.

In the next section we investigate more fully the notion of plu-tori, and the related cosets (weakly special subvarieties) with their underlying dimension notion. This enables us to give a more refined version of Schanuel's conjecture for z^i in Section 3. The uniform version and corresponding Zilber–Pink-type statement are set out in Section 4, where we arrive at the formulation of Theorem 1.3. The subsequent sections are devoted to proving Theorem 1.3 and related statements. We gather some Ax–Schanuel-type statements in Section 5, and then finally in Sections 6–9 we gather the ingredients required to prove Theorem 1.3, first for $V/\overline{\mathbb{Q}}$ and then in a uniform version for families of subvarieties, from which the general case follows.

It has been my privilege over many years now to have Boris Zilber as a colleague, to discuss mathematics with him, and in particular to hear at first-hand his unique and inspirational approach to mathematical structures. I dedicate this paper to Boris and look forward to many further conversations.

2. Plu-tori

We introduce some dimension notions involving pairs of linear relations. We need this notion in the first instance for pairs $(z, w) \in \mathbb{C}^2$, but we need it also for coordinate functions defining linear spaces.

Let

$$D_n = \{(z_1, \dots, z_n, iz_1, \dots, iz_n) : z_1, \dots, z_n \in \mathbb{C}\} \subset \mathbb{C}^n \times \mathbb{C}^n,$$

and let

$$\exp : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$$

be the coordinatewise exponential map. Then $\Gamma_n = \exp(D_n)$.

Definition 2.1. Let V be a finite-dimensional complex vector space. A finite set $(z_1, w_1), \dots, (z_n, w_n)$ of elements of V^2 is called *plu-linearly independent* if they do not satisfy any pair of nontrivial $\mathbb{Q}$-linear equations of the form

$$\sum_j q_j z_j + \sum_j r_j w_j = 0, \tag{1}$$

$$-\sum_j r_j z_j + \sum_j q_j w_j = 0. \tag{1^i}$$

Nontrivial means that $\sum (q_j^2 + r_j^2) \neq 0$. If they do satisfy such a "plu-pair" of equations they are called *plu-linearly dependent*.

Let $B = \{(z_1, w_1), \ldots, (z_n, w_n)\}$. We say that (z_0, w_0) is plu-linearly dependent on B if there is a plu-pair

$$q_0 z_0 + \sum_j q_j z_j + \sum_j r_j w_j + r_0 w_0 = 0,$$
$$-r_0 z_0 - \sum_j r_j z_j + \sum_j q_j w_j + q_0 w_0 = 0,$$

in which q_0 or r_0 is nonzero.

If say $q_1 \neq 0$ in the plu-pair (1), (1^i), then as above we may use (1) to eliminate z_1 but not w_1 from (1^i), and we may use (1^i) to eliminate w_1 but not z_1 from (1) to get a new plu-pair of equations

$$\sum_j (r_1 q_j - q_1 r_j) z_j + \sum_j (q_1 q_j + r_1 r_j) w_j = 0, \qquad (r_1(1) + q_1(1^i))$$
$$\sum_j (-q_1 q_j - r_1 r_j) z_j + \sum_j (r_1 q_j - q_1 r_j) w_j = 0. \qquad (-q_1(1) + r_1(1^i))$$

We now show that plu-linear dependence leads to a well-defined dimension: the cardinality of a maximal plu-independent subset, which we call a *plu-basis*. For this we of course need the exchange property.

Proposition 2.2. *Let B be as above. If (z_0, w_0) is plu-dependent on B and (z_*, w_*) is plu-dependent on $B \cup \{(z_0, w_0)\}$ then (z_*, w_*) is plu-dependent on B.*

Proof. We can assume that the plu-pair for the dependence of (z_0, w_0) on B has the form

$$z_0 + \sum_j q_j z_j + \sum_j r_j w_j = 0, \quad -\sum_j r_j z_j + \sum_j q_j w_j + w_0 = 0.$$

We use these to eliminate z_0, w_0 from the dependence of (z_*, w_*) on $B \cup \{(z_0, w_0)\}$, which remains a plu-pair. □

Proposition 2.3. *Any two plu-bases have the same cardinality.*

Proof. Let B, B' be two maximal plu-linearly independent subsets. If $B = B'$ we are done; otherwise, say $(z_i, w_i) \in B' \backslash B$. By the maximality, (z_i, w_i) is plu-linearly dependent over B. But since B' is plu-linearly independent the plu-pair must have a nonzero coefficient for some $(z_j, w_j) \in B \backslash B'$.

The claim is that B^* with (z_i, w_i) replacing (z_j, w_j) in B is again a maximal plu-linearly independent subset. First, it is plu-linearly independent. Otherwise, we have (z_i, w_i) plu-linearly dependent on $B \backslash \{(z_j, w_j)\}$. But then by Definition 2.1 we would have (z_j, w_j) dependent on $B \backslash \{(z_j, w_j)\}$, a contradiction. But also by Definition 2.1 we see that it "spans".

This shows that $\#B \geq \#B'$. We symmetrically get the other inequality. □

Definition 2.4. A $\mathbb{Q}$-linear subspace $L \subset \mathbb{C}^n \times \mathbb{C}^n$ is called *plu-linear* if the set of $\mathbb{Q}$-linear forms defining it is closed under the operation $(q, r) \mapsto (-r, q)$. A plu-linear $\mathbb{Q}$-subspace is also called a *plu-subspace*.

We observe that plu-linear subspaces have even dimension as linear subspaces. Let L be a plu-subspace. We consider the complex vector space of pairs of complex linear forms

$$\Bigl(\sum_j c_j z_j, \sum_j d_j w_j\Bigr), \quad c_j, d_j \in \mathbb{C},$$

as functions on L. If the coordinate functions z_j, w_j are plu-linearly independent as functions on L then there are no equations and $L = \mathbb{C}^n \times \mathbb{C}^n$. Otherwise, we have a basis of some dimension m and then as a $\mathbb{Q}$-subspace we have $\dim L = 2m$.

The intersection of two (or more) plu-subspaces is a plu-subspace. If $A \subset \mathbb{C}^n \times \mathbb{C}^n$ then there is a smallest plu-subspace containing A, denoted $\langle\!\langle A \rangle\!\rangle$.

If $A \subset D_n$ then the smallest $\mathbb{Q}$-linear subspace of $\mathbb{C}^n \times \mathbb{C}^n$ containing A is a plu-subspace, because the "conjugate" of a given equation follows from multiplying it through by i.

A plu-subspace of dimension $2m$ (as a $\mathbb{Q}$-linear subspace) intersects D_n in a $\mathbb{Q}$-subspace of D_n of dimension at least dimension m since (as the "conjugate" of any given equation holds automatically) the intersection is equal to the intersection of D_n with a $\mathbb{Q}$-subspace defined by $2n - (n + m) = n - m$ independent linear equations, whence has dimension at least $n + (n + m) - 2n = m$. But it is also at most this dimension, as each such equation (with its "conjugate") eliminates one variable. Thus, the intersection of D_n with a plu-subspace of dimension $2m$ is a $\mathbb{Q}$-subspace of D_n of dimension m.

Definition 2.5. A torus $T \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ is called a *plu-torus* if it is the image under exp of some plu-linear $\mathbb{Q}$-subspace $L \subset \mathbb{C}^n \times \mathbb{C}^n$.

The set of exponent vectors $(q, r) \in \mathbb{Z}^n \times \mathbb{Z}^n$ defining a plu-torus is closed under the operation $(q, r) \mapsto (-r, q)$; this is an equivalent condition to the one in Definition 1.1. Each plu-torus $T = \exp(L)$, where L is a plu-subspace of dimension $2m$, contains the image of $\exp(L \cap D_n)$, which we denote $\Gamma_T = \Gamma_n \cap T$.

The intersection of tori is not in general a torus. However, the intersection of two tori contains among its components a unique torus. And if the two tori are plu-tori so is the torus component of their intersection.

Consider a subvariety (i.e., irreducible) $A \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$. Suppose that $A \cap \Gamma_n \neq \varnothing$ and that $A \subset T$ and $A \subset T'$, where T, T' are plu-tori. Then $A \subset T \cap T'$, and hence is contained in one of its components. These are disjoint from each other. The unique preimage of $(x, y) \in A \cap \Gamma_n$ is in D_n and lies in the intersection of the plu-linear subspaces L, L' corresponding to T, T'. We thus see that A is contained in the unique plu-torus component of the intersection.

Thus for $A \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ with $A \cap \Gamma_n \neq \varnothing$ there is a unique smallest plu-torus containing A, which we have denoted $((A))$. In particular, for $(x, y) \in \Gamma_n$ there is a smallest plu-torus $((x, y))$ containing (x, y). And plu-tori have even dimension (as complex subvarieties).

Remark 2.6. More generally, if X is a quasiprojective variety (or even more generally a connected open semialgebraic subset of one; see [Pila 2022, Chapter 14]) we can define a *designated collection* on X to be a collection $\mathcal{S}$ of subvarieties (relatively closed and irreducible) of X such that (i) $X \in \mathcal{S}$ and (ii) $\mathcal{S}$ is closed under taking irreducible components of intersections. This is somewhat more general than the notion of "prespecial structure" considered in [Klingler et al. 2018] (we do not insist that special points be Zariski-dense in special subvarieties), and the still more elaborate setting of "distinguished categories" [Barroero and Dill 2021], but still gives a well-defined notion of "smallest special subvariety containing A" for any $A \subset X$. If $\Omega \subset X$ is some complex analytic subset then one can consider a *designated collection on X meeting* Ω to be a collection of subvarieties of X which have nonempty intersection with Ω, such that (i) $X \in \mathcal{S}$ and (ii) if $Y, Z \in \mathcal{S}$ and W is a component of $Y \cap Z$ which meets Ω then $W \in \mathcal{S}$. Then, as above, one has a well-defined "smallest special subvariety" containing A for any $A \subset X$ which has a nonempty intersection with Ω. This notion arises, as here, naturally in considering ZP-type formulations relevant to certain Schanuel-type statements.

We also want the corresponding "weakly special subvarieties". These come from considering pairs of linear equations modulo some suitable constants.

Definition 2.7. Let V be a finite-dimensional complex vector space. A finite set $(z_1, w_1), \ldots, (z_n, w_n)$ of elements of V^2 is called *strictly plu-linearly independent modulo* $\mathbb{C}$ if there is no nontrivial pair of equations

$$\sum_j q_j z_j + \sum_j r_j w_j = c, \quad -\sum_j r_j z_j + \sum_j q_j w_j = ic$$

with $q_j, r_j \in \mathbb{Q}$ (not all zero), and $c \in \mathbb{C}$.

There is a well-defined notion of *strict plu-mod* $\mathbb{C}$ *basis*, the cardinality of a maximal strictly plu-linearly independent modulo $\mathbb{C}$ subset.

Definition 2.8. A linear subvariety $L \subset \mathbb{C}^n \times \mathbb{C}^n$ is called a *strict plu-linear subvariety* if it is defined by linear equations which are closed under the operation

$$\sum_j q_j z_j + \sum_j r_j w_j = c \mapsto -\sum_j r_j z_j + \sum_j q_j w_j = ic.$$

A strict plu-linear subvariety L has even dimension $2m$. It intersects D_n in a subspace of dimension at least m.

The intersection of two strict plu-linear subspaces is a strict plu-linear subspace. Given $A \subset \mathbb{C}^n \times \mathbb{C}^n$, there is a smallest strict plu-linear subvariety containing it, which we denote $\langle\!\langle A \rangle\!\rangle_{\text{SPL}}$.

Definition 2.9. A torus coset $T \subset \mathbb{G}_{\text{m}}^n \times \mathbb{G}_{\text{m}}^n$ is called a *strict plu-coset* if it is the image under exp of a strict plu-linear subvariety.

Equivalently, a strict plu-coset is a coset T of a torus defined by equations with the property that if $x^q y^r = c$ on T then $x^{-r} y^q = d$ on T for some d with $(c, d) \in \Gamma$.

Consider a subvariety $A \subset \mathbb{G}_{\text{m}}^n \times \mathbb{G}_{\text{m}}^n$. Suppose that $A \cap \Gamma_n \neq \varnothing$ and that $A \subset T$ and $A \subset T'$, where T, T' are strict plu-cosets. Then $A \subset T \cap T'$, and hence is contained in one of its components. These are disjoint from each other. The unique preimage of $(x, y) \in A \cap \Gamma_n$ is in D_n and lies in the intersection of the strict plu-linear subvarieties L, L' corresponding to T, T'. We thus see that A is contained in the unique strict plu-coset component of the intersection.

Thus for $A \subset \mathbb{G}_{\text{m}}^n \times \mathbb{G}_{\text{m}}^n$ with $A \cap \Gamma_n \neq \varnothing$ there is a unique smallest strict plu-coset containing A, which we denote

$$((A))_{\text{SPC}}.$$

There is also a weaker notion of "plu-linear dependence modulo $\mathbb{C}$" in which the pair of constants do not need to be related by multiplication by i. There are corresponding "plu-linear subvarieties" and "plu-cosets", their images under exp.

For $A \subset \mathbb{G}_{\text{m}}^n \times \mathbb{G}_{\text{m}}^n$ there is a unique smallest plu-coset containing A, denoted

$$((A))_{\text{PC}}.$$

3. Schanuel's conjecture for z^i

We can now state a more precise analogue of Schanuel's conjecture for z^i.

Conjecture 3.1 (Schanuel's conjecture for z^i (z^iSC)). Suppose that $(x_i, y_i) \in \Gamma$, $i = 1, \dots, n$. Then

$$\text{t.d.}(x_1, \dots, x_n, y_1, \dots, y_n) \geq \tfrac{1}{2} \dim ((x_1, \dots, x_n, y_1, \dots, y_n)).$$

Here and throughout, "dim" denotes the complex dimension of an algebraic variety. For $n = 1$ the statement reduces to the Gelfond–Schneider theorem and so is true: for if tr.deg.$(x, y) = 0$ we must have $x^i = y$ with $x, y \in \overline{\mathbb{Q}}^\times$. But this is impossible unless $x = 1$ by Gelfond–Schneider, and since we then also have $y^i = 1/x$ we must have $y = 1$ as well, and then they are skew-multiplicatively dependent and $\dim ((x, y)) = 0$.

Remarks 3.2. (1) As a reduct of complex exponentiation, $(\mathbb{C}; +, \times, 0, 1, \Gamma)$ is conjecturally "tame" [Zilber 2005]; in unpublished work, Wilkie has proved it is quasiminimal. (Quasiminimality of the corresponding structure including predicates for *all* complex powers has recently been proven in [Gallinaro and Kirby 2023].)

As an expansion of the real field the structure $(\mathbb{R}, +, \times, \Gamma)$ is also tame, though "d-minimal" (not o-minimal); see [Miller 2005].

(2) Another approach to formulating SC for z^i is to use an equivalent formulation of SC in terms of $q(z) = \exp(2\pi i z)$ rather than $\exp(z)$.

Observe that if $x_1, \dots, x_n$ are multiplicatively independent algebraic numbers then, under SC, their logarithms $u_1, \dots, u_n$, under any determination, are algebraically independent. Hence these u_j, iu_j are certainly linearly independent over $\mathbb{Q}$, and we get a conjectural analogue of Lindemann's theorem (often called the Lindemann–Weierstrass theorem).

Conjecture 3.3 (z^i-Lindemann–Weierstrass conjecture (z^iLW)). Suppose that the algebraic numbers $x_1, \dots, x_n \in \overline{\mathbb{Q}}^\times$ are multiplicatively independent and that $(x_j, y_j) \in \Gamma$, $j = 1, \dots, n$. Then $y_1, \dots, y_n$ are algebraically independent.

In fact this statement already follows from z^iSC, which would seem much weaker than SC.

Proposition 3.4. *z^iSC implies z^iLW.*

Proof. Assume z^iSC. Suppose $x_1, \dots, x_n$ are algebraic and multiplicatively independent. Then their logarithms (under any determination) are linearly independent over $\mathbb{Q}$. Then, by Baker's theorem [Baker 1975], they are linearly independent over $\overline{\mathbb{Q}}$. Then x_j, y_j are plu-multiplicatively independent and so, by z^iSC, $\text{t.d.}(x_i, y_i) = n$. Thus $y_1, \dots, y_n$ are algebraically independent. □

Remark 3.5. Note that if $(z, w) \in \Gamma$ then also $(w, z^{-1}) \in \Gamma$. Thus the analogue of "algebraic independence of logarithms" for z^i, which we might call z^iAIL, is in fact equivalent to z^iLW: *if $y_1, \dots, y_n$ are algebraic and multiplicatively independent and $(x_j, y_j) \in \Gamma$, $j = 1, \dots, n$, then $x_1, \dots, x_n$ are algebraically independent.* It seems interesting to consider other "z^i analogues" of consequences of SC.

4. Uniform Schanuel conjecture and Zilber–Pink conjecture for z^i

We can rephrase z^iSC as follows:

> *Let $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ be defined over $\overline{\mathbb{Q}}$ and with* $\dim V < n$. *If $(x, y) \in V \cap \Gamma$ then (x, y) are plu-multiplicatively dependent.*

More generally, if $T \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ is a $2m$-dimensional plu-torus, $V \subset T$ is defined over $\overline{\mathbb{Q}}$ with $\dim V < m$, and $(x, y) \in V \cap \Gamma$, then (x, y) belongs to a proper plu-subtorus of T.

The uniform version, following [Zilber 2002], asserts that, given such T and V, *finitely many* proper plu-subtori of T account for all such (x, y).

Conjecture 4.1 (uniform Schanuel conjecture for z^i (z^iUSC)). Let $T \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ be a plu-torus of dimension $2m$ and $V \subset T$ an algebraic subvariety, defined over $\overline{\mathbb{Q}}$ with $\dim V < m$. There is a finite set $\mathcal{U}$ of proper plu-subtori of T such that if $(x, y) \in V \cap \Gamma_T$ then $(x, y) \in U$ for some $U \in \mathcal{U}$.

Now let $L \subset \mathbb{C}^n \times \mathbb{C}^n$ be the $\mathbb{Q}$ subspace associated to T. We have $W = (D_n \cap L) \times V \subset L \times T$ of dimension $\dim W < 2m$; the ambient $L \times T$ has dimension $4m$. Therefore, any point in the intersection $V \cap \Gamma_T$ is a point in W on the graph of exp restricted to L. If we assume SC (ideologically speaking assuming z^iSC should be enough, but this is unclear) then, as shown in [Zilber 2002], any point $(x, y) \in V \cap \Gamma$ is in an atypical intersection of V with some plu-subtorus. Thus to get from SC to z^iUSC we need the following Zilber–Pink-type statement, in analogy with "CIT" of [Zilber 2002].

We state the conjecture for $V/\mathbb{C}$ and without dimension restrictions although for the purposes of connecting SC and z^iUSC, only $V/\overline{\mathbb{Q}}$, $2 \dim V < \dim T$ is required.

Conjecture 4.2 (z^iZP). Let $T \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ be a plu-torus. Let $V \subset T$. There is a finite set $\mathcal{U}$ of proper plu-subtori of T with the following property. If $S \subset T$ is a plu-subtorus and $A \subset_{\mathrm{cpt}} V \cap S$ is atypical in dimension with $A \cap \Gamma_T \neq \varnothing$, then there exists $U \in \mathcal{U}$ such that $A \subset U$.

Now given $A \subset T$ with $A \cap \Gamma_n \neq \varnothing$, we have seen that there is a smallest plu-torus containing A, denoted $\langle\langle A \rangle\rangle$, and defined the *plu-defect* $\delta(A) = \dim \langle\langle A \rangle\rangle - \dim A$, and corresponding notion of *plu-optimal* subvariety.

As in [Habegger and Pila 2016], Conjecture 4.2 is then formally equivalent to the statement formulated as Theorem 1.3.

Conjecture 4.3. Let $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$. Then there are only finitely many plu-optimal subvarieties of V.

The reason that we can prove this statement, while multiplicative ZP remains seemingly far out of reach, is the following. An atypical intersection $V \cap S$ is typically a point, and any intersection point is atypical provided $\dim V + \dim S < \dim T$. In [Habegger and Pila 2016] it is shown that the full Zilber–Pink conjecture (in the modular and abelian settings) reduces to finiteness of optimal points in general. Now if we consider "plu-optimal points" they must be algebraic, on the one hand, if $V/\overline{\mathbb{Q}}$, since tori are defined over $\overline{\mathbb{Q}}$, but they must also belong to Γ. But the Gelfond–Schneider theorem (see, e.g., [Baker 1975]) implies that the only such point is $(\mathbf{1}, \mathbf{1})$.

For example, consider the case of a curve $V \subset T$, defined over $\overline{\mathbb{Q}}$. Suppose S is a plu-subtorus and $(x, y) \in V \cap S$. If (x, y) is an isolated intersection then it is algebraic, and consequently $(x, y) \in \Gamma_n$ if and only if $(x, y) = (\mathbf{1}, \mathbf{1})$. So if V intersects atypically in a component meeting Γ_n then this component is either $(\mathbf{1}, \mathbf{1})$

or all of V, in which case V is contained in a proper plu-subtorus. Thus we see that Conjecture 4.3 holds for V.

Our strategy, following [Habegger and Pila 2016], is to apply Ax–Schanuel to reduce to looking for plu-optimal points in the translate spaces of finitely many families, and then the above argument is decisive in showing that there is at most one plu-optimal point, namely the one corresponding to $(\mathbf{1},\mathbf{1})$, in each such family. We go from $V/\overline{\mathbb{Q}}$ to $V/\mathbb{C}$ via a uniform version. The context for this argument is described further in Section 9 where it is presented.

Implementing this strategy requires the analogous "optimal" notions for strict and general plu-cosets. The analogous notion for weakly special subvarieties in the multiplicative setting is "geodesic optimal" (see [Habegger and Pila 2016]), which appeared earlier in [Poizat 2001], and then elsewhere, as "cd-maximal". We define the corresponding defects:

$$\delta_{\mathrm{SPC}}(A)=\dim\,((A))_{\mathrm{SPC}}-\dim A,\qquad \delta_{\mathrm{PC}}(A)=\dim\,((A))_{\mathrm{PC}}-\dim A.$$

Definition 4.4. Let $V\subset\mathbb{G}_{\mathrm{m}}^n\times\mathbb{G}_{\mathrm{m}}^n$. We say that $A\subset V$ is *strictly plu-geodesic optimal* for V if it is maximal for δ_{SPC} among subvarieties of V containing A and meeting Γ, and *plu-geodesic optimal* if it is maximal for δ_{PC} among subvarieties of V containing A.

5. Ax–Schanuel for z^i

Let K be a differential field with $\mathbb{Q}\subset\mathbb{Q}(i)\subset C\subset K$ with commuting derivations D_j and constant field C. The following is a special case of "Ax–Schanuel" [Ax 1971, Theorem 3].

Proposition 5.1. *Let $u_1,\dots,u_n,x_1,\dots,x_n,y_1,\dots,y_n\in K^\times$ with*

$$D_jx_k=x_kD_ju_k,\qquad D_jy_k=iy_kD_ju_k\quad\textit{for all } j,k.$$

Then

$$\mathrm{tr.deg.}_C(u_1,\dots,u_n,x_1,\dots,x_n,y_1,\dots,y_n)\ge 2n+\mathrm{rank}_K(D_ju_k)$$

unless the u_k, iu_k are linearly dependent over $\mathbb{Q}$ modulo C.

Suppose that the u_j, iu_j are linearly dependent over $\mathbb{Q}$ modulo C, say

$$\sum_j q_ju_j+i\sum_j r_ju_j=c\in C \tag{2}$$

with $q_j,r_j\in\mathbb{Z}$ not all zero. Then we find that

$$\prod_j x_j^{q_j}\prod_j y_j^{r_j}=c'\in C,$$

as it is in the kernel of all the derivations. Multiplying (2) through by i we get a second relation (1^i), and a second multiplicative relation

$$\prod_j x_j^{-r_j} \prod_j y_j^{q_j} = c'' \in C.$$

Now morally one wants to say that $c' = \exp(c)$, $c'' = \exp(ic)$ so that $c'' = c'^{\,i}$, but the differential field setting has no interpretation of this.

Conversely, if we are given $x_1, \dots, x_n, y_1, \dots, y_n$ satisfying the differential relations

$$\frac{D_j y_k}{y_k} = i\,\frac{D_j x_k}{x_k} \quad \text{for all } j, k$$

then if $u_1, \dots, u_n$ satisfy $D_j x_k = x_k D_j u_k$ for all j, k then they also satisfy the equations $D_j y_k = i y_k D_j u_k$, and if the x_j, y_j satisfy multiplicative relations mod C then the u_k, iu_k satisfy linear relations over $\mathbb{Q}$ modulo C.

And if we have a linear relation $\sum_j q_j u_j + i \sum_j r_j u_j = c$ then the "conjugate" relation indeed has constant ic.

Finally we note that with any u_j, x_j, y_j as in Proposition 5.1 we have

$$\operatorname{rank}_K(D_j u_k) = \operatorname{rank}_K(D_j x_k) = \operatorname{rank}_K(D_j y_k) = \operatorname{rank}_K(D[x], D[y]),$$

where $D[x] = (D_j x_i)$ and $D[y] = (D_j y_i)$.

Corollary 5.2. *Let* $x_1, \dots, x_n, y_1, \dots, y_n \in K^\times$ *with*

$$\frac{D_j y_k}{y_k} = i\,\frac{D_j x_k}{x_k} \quad \textit{for all } j, k.$$

Then

$$\operatorname{tr.deg.}_C(x_1, \dots, x_n, y_1, \dots, y_n) \geq \tfrac{1}{2} \dim((x_1, \dots, x_n, y_1, \dots, y_n))_{\mathrm{PC}} + \operatorname{rank}_K(D_j x_k),$$

where $((\dots))_{\mathrm{PC}}$ *is defined using plu-multiplicative dependence modulo* C.

Proof. Given x_j, y_j, $j = 1, \dots, n$ satisfying these equations then some suitable u_j exist (perhaps in some extension differential field), and then the statement follows from the above discussion. □

We next state "weak CIT" in this setting, following [Zilber 2002]. Given a variety (or family of varieties) V then there is some finite set of linear dependencies which, up to translations, accounts for all deficiencies in transcendence degree.

Suppose that $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{C}^k$ is a family of algebraic varieties of generic dimension k, parameterized by $t \in W \subset \mathbb{C}^k$, with the fibre having $\dim V_t = k$ provided $t \notin W'$, where W' is a proper subvariety of W.

Proposition 5.3. *Let $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n \times \mathbb{C}^m$ be a family of algebraic varieties, parameterized by points of $W \subset \mathbb{C}^m$, of generic dimension k outside W'.*

Then there exists a finite set Σ of integer vectors $(q, r) \in \mathbb{Z}^{2n} \backslash \{0\}$ with the following property. Suppose $t \in W \backslash W'$, and $A \subset_{\mathrm{cpt}} V_t \cap \Gamma$ with

$$\dim A > k - n.$$

Then there exists $(q, r) \in \Sigma$ and $c, c' \in \mathbb{C}$ such that

$$\prod_j x_j^{q_j} \prod_j y_j^{r_j} = c \quad \textit{and} \quad \prod_j x_j^{-r_j} \prod_j y_j^{q_j} = c'$$

for any point $(x, y) \in A$.

Proof. This is essentially a special case of Proposition 8 of [Zilber 2002], though we consider families of general dimension, not necessarily $k < n$. It is an application of the compactness theorem of first-order logic. Suppose, towards a contradiction, that no such finite set exists. Certainly if $\dim A = 0$ then this property is satisfied for any such tuple. Therefore, for some $\ell > 0$ we have the property that for any finite set Σ of tuples there exists $t \in W \backslash W'$ and a component $A \subset_{\mathrm{cpt}} V_t \cap \Gamma$ of dimension ℓ with $\ell > k - n$. Then it is consistent to have a differential field with ℓ derivations, of rank ℓ on some set $\{x_1, \ldots, x_n\}$ of functions and to have also functions $y_1, \ldots, y_n$ satisfying the required equations but with the x_j, y_j, $j = 1, \ldots, n$ not plu-multiplicatively dependent modulo constants, giving the contradiction. □

By repeating this on the families of intersections (the parameter now being the constants $c', c'' \in \mathbb{C}$), we find that some finite collection of families of plu-cosets accounts for all plu-geodesic optimal intersections with varieties in the family V.

6. The defect condition

Given a plu-torus T and a subvariety $A \subset T$ meeting Γ_T we have three defects: the first with respect to plu-tori, the second with respect to strict plu-cosets, and the third with respect to (general) plu-cosets:

$$\delta(A) = \dim\,((A)) - \dim A,$$

$$\delta_{\mathrm{SPC}}(A) = \dim\,((A))_{\mathrm{SPC}} - \dim A,$$

$$\delta_{\mathrm{PC}}(A) = \dim\,((A))_{\mathrm{PC}} - \dim A.$$

Evidently

$$\delta_{\mathrm{PC}} \leq \delta_{\mathrm{SPC}}(A) \leq \delta(A).$$

Suppose that (x_i, y_i) are a basis of coordinate pairs on A with respect to plu-multiplicative dependence. The difference between the defect measures to what extent these functions are strictly plu-dependent mod $\mathbb{C}$, and then the extent to which the remaining strictly plu-dependent mod $\mathbb{C}$ ones are plu-multiplicatively dependent mod $\mathbb{C}$.

Suppose $A \subset B$. Then strict plu-multiplicative relations modulo constants on B remain strict plu-multiplicative relations modulo constants on A, and if the constant pairs (c_k, d_k) in these relations are plu-multiplicatively independent on B they remain so on A. We therefore see that the *defect condition* holds between δ and δ_{SPC}, namely

$$\delta(B) - \delta_{\mathrm{SPC}}(B) \leq \delta(A) - \delta_{\mathrm{SPC}}(A).$$

Similarly, the defect condition holds between δ_{PC} and δ_{SPC}.

Proposition 6.1. *Fix $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$.*

(i) *A plu-optimal subvariety for V is strictly plu-geodesic optimal.*

(ii) *A strictly plu-geodesic optimal subvariety is plu-geodesically optimal.*

Proof. Suppose $A \subset V$ is plu-optimal and $A \subset B \subset V$. Suppose $\delta_{\mathrm{SPC}}(B) \leq \delta_{\mathrm{SPC}}(A)$. Then, by the defect condition,

$$\delta(B) \leq \delta(A) - \delta_{\mathrm{SPC}}(A) + \delta_{\mathrm{SPC}}(B) \leq \delta(A).$$

Since A is optimal, we have $B = A$, and so A is strictly plu-geodesically optimal. This proves (i). The proof of (ii) is similar. □

7. Families of plu-cosets

In this section we introduce some terminology and notation that will be needed in the proofs of the main results. For general properties of linear tori see [Bombieri and Gubler 2006, Chapter 3.1]. A family of (general) plu-cosets of a plu-torus X is determined by a finite set of plu-multiplicative pairs of equations

$$x^{q^{(k)}} y^{r^{(k)}} = 1, \quad x^{-r^{(k)}} y^{q^{(k)}} = 1 \quad \text{for } k = 1, \dots, K$$

which define X, and a finite set of exponent vectors, independent of those above, for some further equations which determine the cosets in the family:

$$x^{q^{(k)}} y^{r^{(k)}} = c_k, \quad x^{-r^{(k)}} y^{q^{(k)}} = d_k \quad \text{for } k = K+1, \dots, K+L,$$

such that the fibres are torus cosets in X (i.e., the exponent vectors generate a primitive lattice).

The plu-cosets are parameterized by the coordinates

$$(c, d) = (c_{K+1}, \dots, c_{K+L}, d_{K+1}, \dots, d_{K+L}) \in \mathbb{G}_{\mathrm{m}}^L \times \mathbb{G}_{\mathrm{m}}^L.$$

We denote the parameter space by $X_S = \mathbb{G}_{\mathrm{m}}^L \times \mathbb{G}_{\mathrm{m}}^L$. The cosets are then the fibres of the map $\pi : X \to X_S$ given by

$$\pi(x, y) = \big(x^{q^{(K+1)}} y^{r^{(K+1)}}, \dots, x^{q^{(K+L)}} y^{r^{(K+L)}}, x^{-r^{(K+1)}} y^{q^{(K+1)}}, \dots, x^{-r^{(K+L)}} y^{q^{(K+L)}}\big).$$

Such a family we denote S^D, where D is the data (exponent vectors for the equations of T and for the additional equations of the cosets in the family). The fibre over $(c, d) \in X_S$ is denoted $S^D_{c,d}$. The fibre $S^D_{c,d}$ is a strict plu-coset just if $(c, d) \in \Gamma_L$. The union X of the plu-cosets over the family we call the *envelope* of S^D and denote it $[S^D]$.

We observe that the preimage under π of a plu-torus is a plu-torus, and likewise for strict and general plu-cosets. To see this, consider a condition of the form

$$c^Q d^R = \gamma, \quad c^{-R} d^Q = \delta,$$

where $Q = (Q_1, \dots, Q_L)$, $R = (R_1, \dots, R_L)$ are tuples of integers. The preimage in X is determined, in addition to the equations for X, by

$$x^{\sum Q_j q^{(j)} - \sum R_j r^{(j)}} y^{\sum Q_j r^{(j)} + \sum R_j q^{(j)}} = \gamma,$$
$$x^{-\sum R_j q^{(j)} - \sum Q_j r^{(j)}} y^{-\sum R_j r^{(j)} + \sum Q_j q^{(j)}} = \delta,$$

which is a plu-pair.

8. Proof of Theorem 1.3 for $V/\overline{\mathbb{Q}}$

We can now prove Theorem 1.3 (Conjecture 4.3) for $V/\overline{\mathbb{Q}}$ following the first part of the proof of [Habegger and Pila 2016, Theorem 10.1], using the fact that Γ has just one algebraic point (a consequence of the Gelfond–Schneider theorem).

Theorem 8.1. *Let $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$ with $V/\overline{\mathbb{Q}}$. Then there are only finitely many plu-optimal subvarieties of V.*

Proof. Let $A \subset V$ be a plu-optimal component which meets Γ_n.

We observe that if $A \subset S$ for some plu-coset S then this coset must be strict. For suppose $(x, y) \in A \cap \Gamma_n$, so $(x, y) \in S$. Let (u, iu) be the tuple of logarithms. Say that $x^q y^r = c$, $x^{-r} y^q = d$ is a pair of equations defining S. So we have

$$\sum q_j u_j + i \sum r_j u_j = \gamma$$

with $\exp(\gamma) = c$. But then, multiplying by i, we also have

$$-\sum r_j u_j + i \sum q_j u_j = i\gamma,$$

whence $d = \exp(i\gamma)$.

Plu-optimal subvarieties are plu-geodesic optimal. The plu-geodesic-optimal subvarieties of V arise from intersections with finitely many families S^D of plu-cosets. Fix one of these families S and denote the parameter space X_S. Let $\tau = \dim S_{c,d}$ be the dimension of the plu-cosets in the family.

We have a projection $\pi : [S] \to X_S$ whose fibres are the $S_{c,d}$. The intersections $V \cap S_{c,d}$ are the fibres of the restriction of this projection to V, whose image in X_S we denote V_S, and we have $V_S/\overline{\mathbb{Q}}$. There is a Zariski-open subvariety $V' \subset V$ in which the fibre over the image has the generic fibre dimension $\nu = \dim V - \dim V_S$.

Suppose $A \subset V \cap S_{c,d}$ is plu-optimal and meets Γ_n. If $A \subset V \backslash V'$ then it is certainly plu-optimal for the component of $V \backslash V'$ it is in. The proof is then concluded by induction on $\dim V$ (the base case $\dim V = 1$ was dealt with in the second paragraph following the statement of Conjecture 4.3).

So we assume that $A \cap V' \neq \varnothing$ and then $\dim A = \nu$ and $\pi(A) = (c, d)$. Since $A \cap \Gamma_n \neq \varnothing$ we have some $(x, y) \in A \cap \Gamma_n$. Thus each component $(x_i, y_i) \in \Gamma_1$, with logarithm u_i. And now if

$$\sum_j q_j u_j + i \sum_j r_j u_j = \gamma$$

then this implies

$$-\sum_j r_j u_j + i \sum_j q_j u_j = i\gamma.$$

Then $c = \exp(\gamma)$, $d = \exp(i\gamma)$ and we have $(c, d) \in \Gamma$.

The claim is that $\{(c, d)\}$ is a plu-optimal point component of V_S. We can assume that we have already dealt with any family of smaller plu-cosets that might have given rise to A, i.e., we can assume that

$$S_{c,d} = ((A))_{\mathrm{PC}}.$$

Then

$$\dim((A)) = \dim S_{c,d} + \dim((c, d)),$$

whence

$$\delta(A) = \dim((A)) - \dim A = \dim((c, d)) + \tau - \nu.$$

Suppose that $\{(c, d)\} \subset B$, $\{(c, d)\} \neq B$ with $\delta(B) \le \delta(c, d) = \dim((c, d))$. Let C be the component of the preimage of B in V' containing A. Then

$$\delta(C) = \dim((C)) - \dim C \le \dim((B)) + \tau - (\dim B + \nu) \le \dim((c, d)) + \tau - \nu = \delta(A).$$

Then $C = A$ by the plu-optimality of A and so $B = \{(c, d)\}$ is optimal.

But then (c, d) is algebraic, and since it belongs to Γ_n we must have $(c, d) = (\mathbf{1}, \mathbf{1})$. So we get at most one plu-optimal subvariety in each family. □

Remarks 8.2. (1) Can one effectively determine the finitely many families of plu-cosets? (For the general multiplicative setting [Bombieri et al. 2007] gives an effective argument for this.) For curves this seems clearly possible.

(2) This shows that SC would imply a uniform z^iSC. Does z^iSC itself imply some uniformity, and if so what is the intervening "ZP" statement?

9. Uniformity and proof of Theorem 1.3 for $V/\mathbb{C}$

Here we prove that Conjecture 4.3 holds uniformly for varieties in families, in the sense of [Scanlon 2004]: the formal sum of the optimal subvarieties is bounded as a cycle, which we make precise in Conjecture 9.2. This uses the fact established in [Habegger and Pila 2016], already exploited here, that ZP is equivalent to showing that the number of optimal points on any subvariety is bounded. Here we need to upgrade this to show that the number of optimal points is uniformly bounded on a family of varieties. We do this following the argument sketched in [Zannier 2012], which we have fully worked out [Pila 2022, Chapter 24] for ZP in the modular and multiplicative settings. As a by-product, we establish that Conjecture 4.3 holds for $V/\mathbb{C}$.

Let $X = \mathbb{G}_\mathrm{m}^n \times \mathbb{G}_\mathrm{m}^n$. A *family of subvarieties* of X means a subvariety $V \subset X \times P$ for some constructible set P, considered as the family of fibres $V_p \subset X$, $p \in P$. The *fibre dimension* of a family is the maximum dimension of a fibre.

If V is a family of subvarieties of X and h is a positive integer then we have the incidence variety

$$\mathrm{Inc}^h(V) = \{(z_1, \ldots, z_h) \in X^h : \exists p \in P : z_j \in V_p,\ j = 1, \ldots, h\}.$$

Since P is only assumed constructible $\mathrm{Inc}^h(V)$ may not be Zariski closed, and we denote by $V^{\langle h\rangle}$ its Zariski closure. In particular $V^{\langle 1\rangle}$ is the Zariski closure of the union of all the fibres, which we call the *envelope* of the family and denote also by $[V]$.

We have already seen that plu-cosets of a plu-torus $T \subset X$ come in families. Such a family is a family $S \subset X \times X_S$ (in the above sense). The envelope $[S]$ is a plu-torus.

Theorem 9.1. *Let $V \subseteq X \times P$ be a family of subsets of X defined over $\overline{\mathbb{Q}}$. Then there is a uniform bound on the number of optimal points of V_p, $p \in P$.*

Proof. We prove the theorem by induction (first) on the dimension of the parameter space P. The case $\dim P = 0$ is addressed by Theorem 8.1. We may then assume that the Zariski closure of P is irreducible, that the fibre dimension $v = \dim V_p$ is constant and equal to the generic fibre dimension, and that there is a single family $S = S^D$ of plu-cosets such that every fibre V_p in the family has $((V_p))_{PC} = S_{c,d}$ for some $(c, d) \in X_S$. Thus $[V] \subset [S]$. We adopt the notation of Section 7 for this family.

For given $\dim P$ we may assume that the theorem holds for all families $W \subset X \times Q$ (i.e., of any dimension of the parameter space Q) such that each fibre has $((W_q))_{PC} = S'_{c,d}$ for some fibre of a family S' of plu-cosets, and for which either $\dim[S'] < \dim[S]$ or $\dim[S'] = \dim[S]$ and $\dim S'_{c,d} < \dim S_{c,d}$; the base cases are trivial.

Now we take a positive integer h, to be specified below, and consider $V^{\langle h\rangle} \subset [S]^h$. Then in fact $V^{\langle h\rangle} \subset S^{\langle h\rangle}$, and the latter is a plu-torus: in addition to the equations for $[S]^h$ it is defined by the plu-pairs of equations

$$(x^{(j)})^{q^{(\ell)}}(y^{(j)})^{r^{(\ell)}} = (x^{(k)})^{q^{(\ell)}}(y^{(k)})^{r^{(\ell)}}, \quad (x^{(j)})^{-r^{(\ell)}}(y^{(j)})^{q^{(\ell)}} = (x^{(k)})^{-r^{(\ell)}}(y^{(k)})^{q^{(\ell)}}$$

for $j \neq k$ and $\ell = K+1, \ldots, K+L$. Thus

$$\dim S^{\langle h\rangle} = h \dim[S] - L(h-1).$$

Now suppose that V_p is a fibre of V which contains h optimal points $(x_0^{(j)}, y_0^{(j)})$, $j = 1, \ldots, h$. Then they are atypical as point subvarieties of V_p (unless $\dim V_p = 0$ in which case the conclusion is trivial for V). Thus there are plu-tori T_j, $j = 1, \ldots, h$ such that $(x_0^{(j)}, y_0^{(j)}) \in T_j \cap \Gamma_X$, $j = 1, \ldots, h$ and

$$\dim T_j + \dim V_p < \dim[S].$$

Consider the plu-torus

$$T = T_1 \times \cdots \times T_h.$$

Since the equations defining $S^{\langle h\rangle}$ are between different groups of variables, they are independent of the equations defining each T_j, and we have

$$\dim T \cap S^{\langle h\rangle} = \dim T - L(h-1).$$

Then

$$(\boldsymbol{x}_0, \boldsymbol{y}_0) = \big((x_0^{(1)}, \ldots, x_0^{(h)}, y_0^{(1)}, \ldots, y_0^{(1)})\big) \in T \cap S^{\langle h\rangle} \cap \Gamma_T$$

and is atypical for $V^{\langle h\rangle}$ as a subvariety of $S^{\langle h\rangle}$ provided that

$$\dim V^{\langle h\rangle} + \dim T \cap S^{\langle h\rangle} < \dim S^{\langle h\rangle}.$$

Thus we find that $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ is atypical provided

$$\dim P + h \dim v + h(\dim[S] - v - 1) - L(h-1) < h \dim[S] - L(h-1),$$

that is, provided $h > \dim P$. We now assume this (but the choice of h needs to be on the basis of some combinatorial principles further below).

We can now apply Theorem 8.1 to $V^{\langle h\rangle}$, which is defined over $\overline{\mathbb{Q}}$, to conclude that atypical points are contained in one of finitely many proper plu-subtori $U \subset S^{\langle h\rangle}$. Each such U is determined by at least one plu-pair of equations

$$\prod_j (x^{(j)})^{s^{(j)}}(y^{(j)})^{t^{(j)}} = 1, \quad \prod_j (x^{(j)})^{-t^{(j)}}(y^{(j)})^{s^{(j)}} = 1.$$

Consider the exponent vector pair $(s(\ell), t(\ell))$ on a particular set of variables. If this is not in the lattice $\Lambda(S)$ generated by the equations defining S (the fixed ones and the variable one), and if we have some points $(x^{(j)}, y^{(j)})$, $j \neq \ell$ and sufficiently many points $(x^{(\ell)}, y^{(\ell)})$ such that, for each of the $(x^{(\ell)}, y^{(\ell)})$, the relation

$$(x^{(1)})^{s^{(1)}} \cdots (x^{(\ell)})^{s(\ell)} \cdots (x^{(h)})^{s^{(h)}} (y^{(1)})^{t^{(1)}} \cdots (y^{(\ell)})^{t(\ell)} \cdots (y^{(h)})^{t^{(h)}} = 1,$$

and the companion relation

$$(x^{(1)})^{-t^{(1)}} \cdots (x^{(\ell)})^{-t(\ell)} \cdots (x^{(h)})^{-t^{(h)}} (y^{(1)})^{s^{(1)}} \cdots (y^{(\ell)})^{s(\ell)} \cdots (y^{(h)})^{s^{(h)}} = 1,$$

hold, then we get many points $(x^{(\ell)}, y^{(\ell)}) \in V_p$ satisfying an additional plu-relation mod $\mathbb{C}$ with the exponent-pair $(s(\ell), t(\ell))$. The corresponding family S' of plu-cosets has smaller fibre dimension than S.

If $(s(\ell), t(\ell))$ is in $\Lambda(S)$, then using the relations defining $S^{\langle h\rangle}$ we can replace this plu-pair by an equivalent pair of equations with trivial exponent pair $(s(\ell), t(\ell))$. It may be that, for some U, every $(s(\ell), t(\ell)) \in \Lambda(S)$. But since the equations defining U define a proper plu-subtorus of $S^{\langle h\rangle}$, when the relation is shifted to a single set of variables it must be one that does not hold identically on $[S]$, but gives a proper plu-subtorus S_U.

We thus have that, for any h optimal points on some fibre V_p, we get one of finitely many possibilities: that some designated coordinate lies in one of the S_U, or the h-tuple of tuples satisfies one of finitely many relations involving exponent pairs that, wherever they are nontrivial on a group of variables, do not belong to $\Lambda(S)$.

If we now choose a much larger $\mathcal{H}$ and have plu-optimal $(x_0^{(j)}, y_0^{(j)})$, $j = 1, \ldots, \mathcal{H}$ in some order, then by the hypergraph Ramsey theorem we can be assured that there is some subset of H of them for which all choices (in order) of h satisfy the same one of these conditions. We thus find that we have many points on some family of plu-cosets of smaller dimension, or many points in some smaller plu-torus S_U. We can therefore complete the proof by induction. □

A rephrasing of Conjecture 4.3 is that, for $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n$, the formal sum of plu-optimal subvarieties is a cycle V^{opt}. We now want to frame a uniform version that, over families V, the plu-optimal cycle is uniformly bounded as a cycle. We formulate this following [Scanlon 2004].

It should be borne in mind that z^iZP makes a nontrivial statement only for subvarieties that meet Γ_n. If V does not meet Γ_n then, by definition, V has no plu-optimal subvarieties and V^{opt} is empty, while if V does meet Γ_n then V itself is plu-optimal for V and V^{opt} is nonempty.

Conjecture 9.2 (Uz^iZP). Let $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n \times P$ be a family of subvarieties. Then there is a family $W \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n \times Q$ such that, for every $p \in P$ for which V_p meets Γ_n, there exists $q \in Q$ such that $V_p^{\mathrm{opt}} = W_q$.

Theorem 9.3. *Uz^iZP holds for families V defined over $\overline{\mathbb{Q}}$.*

Proof. We replay the proof of Theorem 8.1. The plu-geodesic optimal subvarieties of all the V_p come in finitely many families. For each such family, the plu-optimal subvarieties on the fibres of V correspond to plu-optimal points on the family of fibres of the projections. For each such family there is a uniformly bounded number of plu-optimal points on a fibre by Theorem 9.1. □

Corollary 9.4 (Theorem 1.3). *Conjecture 4.3 holds for* $V/\mathbb{C}$.

Proof. Every such V is a fibre in a family defined over $\overline{\mathbb{Q}}$. □

In a similar way, Theorem 9.3 holds for families defined over $\mathbb{C}$ as every such family $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n \times P$ is a subfamily (meaning its fibres are a subset of the fibres) of a larger family $V \subset \mathbb{G}_{\mathrm{m}}^n \times \mathbb{G}_{\mathrm{m}}^n \times Q$ defined over $\overline{\mathbb{Q}}$.

The present results generalize suitably to other algebraic powers, as will be shown in forthcoming work of Cassani. It seems also interesting to consider modular analogues.

Acknowledgements

I thank Daniel Bertrand, Oliviero Cassani, Luca Ghidelli, Jonathan Kirby, Alex Wilkie, and Boris Zilber for their comments, clarifications, and suggestions, and I thank the referee for catching several errors and obscurities. I am grateful to the EPSRC for partial support of this research under grant number EP/N008359/1.

References

[Ax 1971] J. Ax, "On Schanuel's conjectures", *Ann. of Math.* (2) **93** (1971), 252–268. MR Zbl

[Baker 1975] A. Baker, *Transcendental number theory*, Cambridge Univ. Press, 1975. MR Zbl

[Barroero and Dill 2021] F. Barroero and G. A. Dill, "Distinguished categories and the Zilber–Pink conjecture", preprint, 2021. arXiv 2103.07422

[Bays and Kirby 2018] M. Bays and J. Kirby, "Pseudo-exponential maps, variants, and quasiminimality", *Algebra Number Theory* **12**:3 (2018), 493–549. MR Zbl

[Bombieri and Gubler 2006] E. Bombieri and W. Gubler, *Heights in Diophantine geometry*, New Math. Monographs **4**, Cambridge Univ. Press, 2006. MR Zbl

[Bombieri et al. 2007] E. Bombieri, D. Masser, and U. Zannier, "Anomalous subvarieties — structure theorems and applications", *Int. Math. Res. Not.* **2007**:19 (2007), art. id. rnm057. MR Zbl

[Caycedo and Zilber 2014] J. D. Caycedo and B. Zilber, "Natural models of theories of green points", preprint, 2014. arXiv 1401.3326

[Gallinaro and Kirby 2023] F. Gallinaro and J. Kirby, "Quasiminimality of complex powers", preprint, 2023. arXiv 2304.06450

[Habegger and Pila 2016] P. Habegger and J. Pila, "O-minimality and certain atypical intersections", *Ann. Sci. Éc. Norm. Supér.* (4) **49**:4 (2016), 813–858. MR Zbl

[Klingler et al. 2018] B. Klingler, E. Ullmo, and A. Yafaev, "Bi-algebraic geometry and the André–Oort conjecture", pp. 319–359 in *Algebraic geometry* (Salt Lake City, 2015), edited by T. de Fernex et al., Proc. Sympos. Pure Math. **97.2**, Amer. Math. Soc., Providence, RI, 2018. MR Zbl

[Miller 2005] C. Miller, "Tameness in expansions of the real field", pp. 281–316 in *Logic Colloquium '01* (Vienna, 2001), edited by M. Baaz et al., Lect. Notes Log. **20**, Assoc. Symbol. Logic, Urbana, IL, 2005. MR Zbl

[Pila 2022] J. Pila, *Point-counting and the Zilber–Pink conjecture*, Cambridge Tracts Math. **228**, Cambridge Univ. Press, 2022. MR Zbl

[Pink 2005] R. Pink, "A common generalization of the conjectures of André–Oort, Manin–Mumford, and Mordell–Lang", preprint, 2005, available at https://people.math.ethz.ch/~pink/publications.html.

[Poizat 2001] B. Poizat, "L'égalité au cube", *J. Symbolic Logic* **66**:4 (2001), 1647–1676. MR Zbl

[Scanlon 2004] T. Scanlon, "Automatic uniformity", *Int. Math. Res. Not.* **2004**:62 (2004), 3317–3326. MR Zbl

[Zannier 2012] U. Zannier, *Some problems of unlikely intersections in arithmetic and geometry*, Ann. Math. Studies **181**, Princeton Univ. Press, 2012. MR Zbl

[Zilber 2002] B. Zilber, "Exponential sums equations and the Schanuel conjecture", *J. London Math. Soc.* (2) **65**:1 (2002), 27–44. MR Zbl

[Zilber 2003a] B. Zilber, "Raising to powers in algebraically closed fields", *J. Math. Log.* **3**:2 (2003), 217–238. MR Zbl

[Zilber 2003b] B. Zilber, "A remarkable pseudo-analytic structure", preprint, 2003, available at https://people.maths.ox.ac.uk/zilber/publ.html.

[Zilber 2005] B. Zilber, "Pseudo-exponentiation on algebraically closed fields of characteristic zero", *Ann. Pure Appl. Logic* **132**:1 (2005), 67–95. MR Zbl

[Zilber 2015] B. Zilber, "The theory of exponential sums", preprint, 2015. arXiv 1501.03297

Received 25 Nov 2022. Revised 23 Feb 2023.

JONATHAN PILA:

pila@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.647

Zilber's notion of logically perfect structure: universal covers

John T. Baldwin and Andrés Villaveces

We sketch the mathematical background and the main ideas in the proofs of categoricity of theories of several examples of universal covers — reducing an analytic to a model theoretic (discrete) description. We hope this discussion will be useful to a wide spectrum of mathematicians, ranging from those working in geometry to those working in logic; specifically, model theory.

1. Introduction 647
2. Model theory in mathematics 652
3. Categoricity, quasiminimality and excellence 655
4. Modular and Shimura curves 659
5. First order excellence 673
6. Higher-dimensional Shimura varieties 676
7. Model theory and analysis 677
8. Families of covers of algebraic curves 678
Acknowledgements 680
References 681

1. Introduction

The goal of this paper is to sketch (hopefully for a wide spectrum of mathematicians ranging from those working in geometry to those working in logic; specifically, model theory) some recent interactions between model theory and a roughly 150-year old study of analytic functions involving complex analysis, algebraic topology, and number theory that explore the canonicity of universal covers. Towards this goal we discuss and present several examples indicating the main ideas of the proofs and the necessary changes in method for different situations.

Here is Zilber's description of his own project (from his 2000 Logic Colloquium talk in Paris [52]):

MSC2020: primary 03C45, 03C75, 03C98, 11U09, 12L12; secondary 03-03.
Keywords: model theory, infinitary logic, model theory of covers, abelian varieties, Shimura varieties, modular invariants, categoricity.

> *The initial hope of this author in [51] that any uncountably categorical structure comes from a classical context (the trichotomy conjecture), was based on the belief that logically perfect structures could not be overlooked in the natural progression of mathematics. Allowing some philosophical license here, this was also a belief in a strong logical predetermination of basic mathematical structures. As a matter of fact, it turned out to be true in many cases. … Another situation where this principle works is the context of o-minimal structures [38].*

A rather ambitious project aimed at finding *categorical* axiomatizations (Definition 3.0.1) of various kinds of *universal covers* has been unfolding in the twenty-first century. The simplest example of such universal covers is given by the short exact sequence

$$0 \to \ker(\exp) \to (\mathbb{C}, +, 0) \xrightarrow{\exp} (\mathbb{C}, +, \cdot, 0, 1) \to 1. \tag{1}$$

Zilber's original project really aimed to understand the sequence

$$0 \to \ker(\exp) \rightarrowtail (\mathbb{C}, +, \cdot, \exp) \xrightarrow{\exp} \!\!\!\!\!\!\twoheadrightarrow (\mathbb{C}, +, \cdot, \exp) \to 1. \tag{2}$$

The first diagram describes a two-sorted *cover* of the multiplicative group by the additive group. The *full* field structure is studied on the range space although the kernel is of the homomorphism from $(\mathbb{C}, +, 0)$ to $(\mathbb{C}, \cdot, 1)$.

The second [54] corresponds to the *theory* of the complex exponential field. The domain and range of the map are the same exponential field but the kernel is again computed with respect to the homomorphism exp from $(\mathbb{C}, +) \to (\mathbb{C}^*, \times)$.

In both cases, first order axioms are supplemented by an $L_{\omega_1,\omega}$-sentence asserting the kernel is isomorphic to $\mathbb{Z}$, i.e., is standard. Here, we focus on three main *families* of generalizations (described in the chart below) of the first diagram. As this question was extended to more general algebraic contexts, the fundamental cover diagram from (1) changed to this more general situation:

$$C \overset{p}{\twoheadrightarrow} S(\mathbb{C}). \tag{3}$$

Notice two things:

- The map p remains a projection, but it will significantly change as the family of examples unfolds.
- There is no longer a kernel when $S(\mathbb{C})$ is not a group.

Therefore, in a Protean way, the infinitary description that in the particular case described a 'standard kernel' assumes various guises for different examples. Usually, the descriptions are of 'standard fibers' rather than having a 'standard kernel'.

Crucially, in all cases except part of Section 5 the target will be some kind of definable set in an algebraically closed field. The necessary vocabulary for the domain will vary among the situations considered. Shimura varieties require a more general domain:

Notation 1.0.1 (the general situation).

$$X^+ \xrightarrow{p} S(\mathbb{C}) \to 1. \tag{4}$$

Here, $S(\mathbb{C})$ is a variety arising as the quotient of the action of a discrete group on $\mathbb{H}$ (hyperbolic space) or more generally (Shimura varieties) on a hermitian symmetric domain X^+. The target is described by a first order theory $T := \mathrm{Th}(S(\mathbb{C}))$ in a large enough (field) countable vocabulary with quantifier elimination (possible, as S is definable in $(\mathbb{C}, +, \times)$). Notation 1.0.1 thus instantiates the general schema, with appropriate notations for specific cases to be given as we discuss them. Zilber describes the value of his project in terms of 'a complete formal invariant' (Remark 5.3.2).

> *The geometric value of the project is perhaps in the fact that the formulation of the categorical theory of the universal cover of a variety X ... is essentially a formulation of a complete formal invariant of X.* [16, §1]

Table 1 organizes the papers which are the major source for this study. It also provides a keyword describing the main method or context used, and the section of this paper where issues around the specific variant are explained.

The first row of the table — an axiomatization of the exponential map from the complex field to itself; see [54] — differs from the others in the role of the quantifier 'there exists uncountably many'. In that case it is essential to directly control the *cardinality of the algebraic closure of a countable set*; moreover, the

topic	section	sources	method/context
complex exponentiation	1	[53]	quasiminimality
cov mult group	1	[55; 6]	quasiminimality
j-function	4.1	[25]	background
modular/Shimura curves	4	[15; 16]	quasiminimality
finite Morley rank groups	5.1	[7]	fmr and notop
abelian varieties	5.3	[9]	fmr and notop; quasiminimality
Shimura varieties	6	[19]	notop
smooth varieties	8	[57]	o-quasiminimality

Table 1. Chief sources for the general topics covered in this study, together with the main method or context used.

domain has a field structure that disappears in the two-sorted approach of the rest. In the remaining rows, the infinitary logic $L_{\omega_1,\omega}$ is used to control the size of fibers of the cover or when the structure is a group the size of the kernel. This requirement suffices to also control the cardinality of the algebraic closure.

The next block of rows deals with *curves* (1-dimensional objects), where categoricity is obtained by quasiminimality.

The following block deals with higher-dimensional varieties; those rows stray from formal categoricity towards more traditional descriptions of models, and quasiminimality is replaced by a different version of excellence arising in Shelah's study of notop theories (an important notion in classification theory). Both quasiminimality and 'notop' apply to abelian varieties.

The last row considers families of covers of arbitrary smooth algebraic varieties with an infinitary logic construction defined over o-minimal expansions of the reals. There, the focus is on categoricity in $\aleph_1$.

It is worth noting that we could have organized our chart under a totally different scheme. The abelian varieties and $(\mathbb{C}, +)$ are specific varieties. The j-function and the Shimura varieties may be regarded as moduli spaces for (generalized) families of varieties.[1] After preliminary discussions on the model theoretic framework, in Section 4 we sketch in some detail categoricity of universal covers of modular curves. In the later sections we describe the modifications to this program necessary for higher dimensions.

Mathematical encounters.

Some ancient history: in and out of the Zilber world. The first author turns to the first person singular for some memories:

Zilber and I both received our Ph.D.'s in the early 1970's. An important result appeared in both theses: the solution to Morley's conjecture that an $\aleph_1$-categorical theory has finite Morley rank. Such an overlap was not an issue during the Cold War. (On the other hand, my advisor, Lachlan, had to write an entirely new thesis when the result of the proposed one appeared in the west as he was about to submit.)

Given my zero knowledge of Russian, I first learned in any detail of Zilber's work during the 1980–81 model theory year in Jerusalem. Greg Cherlin had no such deficiency and gave with Harrington and Lachlan an alternate proof of Zilber's theorem that there were no finitely axiomatizable totally categorical theories. They relied on the classification of finite simple groups. A few years later Boris completed his model theoretic proof of the key combinatorial lemma avoiding that reliance.

I first knew Boris in any depth during the model theory semester in Chicago 1991–92. Unfortunately, I had partially financed a semester by agreeing to be acting

[1] Types of Shimura varieties include Siegel, PEL-type, and Hodge-type; only some parametrize algebraic varieties.

head the Fall semester, thereby restricting my mathematical activity. In that busy fall, Boris and Angus Macintyre lectured on Tuesday's on Zariski geometries and o-minimality, respectively. The lively group include Macintyre, Zilber, Laskowski, Marker, Otero, D'Aquino and myself, with Pillay driving in weekly from Notre Dame. Lunch was at a deli that Boris insisted on because of the soup followed by coffee at Jamoch's, the first modern coffee house in the UIC area.

About that time, I began work on the Hrushovski construction, but in a quite different direction from Boris: predimension with irrational α. This led to my work with Shelah giving the first full proof of the 0-1 law with edge probability $n^{-\alpha}$ and that the theory of the Shelah–Spencer graph was stable, building on the 1992 Ph.D. thesis of my student Shi. And this led to work with Kitty Holland on fusions, giving the first construction of a rank 2 field with a definable infinite predicate. And then back to Boris and his work on complex exponentiation. Understanding his notion of quasiminimal excellence inspired the desire to understand Shelah's more general notion of excellence. Thence came my monograph on abstract elementary classes and subsequent work on infinitary logic. In any case, visits several times a decade to Oxford always were exciting sources of ideas and pleasant times.

An unlikely encounter of two areas: MAMLS at Rutgers, 2001. The second author of this paper witnessed and participated in one of those momentous encounters of two areas that only seldom happen, and recounts it in the first person singular:

During the MAMLS Meeting at Rutgers in February 2001, a group of people working in Abstract Elementary Classes (including Rami Grossberg, Monica VanDieren, Olivier Lessmann and myself) was very busy discussing Shelah's notion of excellence, originally linked to his work in the model theory of $L_{\omega_1,\omega}$. The *n-amalgamation diagram* was very much part of that discussion. There was a lecture by Boris Zilber at the end of the day, and we all attended, not expecting to understand much, but eager to see him speak. To our great surprise, at the end of Zilber's lecture (dealing with exponential covers, mentioning many analytic number theoretic methods that were arcane to us, and mixing in areas such as Nevanlinna theory), he asked a final question and drew a picture underscoring his question. Boris's picture was *exactly* the n-amalgamation diagram we had been discussing thoroughly with the AEC people those very same days; his question was exactly about the behavior of types in the amalgam and how it could be controlled by small pieces in the components. We jumped to talk to him at the end of his lecture, with the excitement of seeing a potential connection. Boris said he didn't know the model theory of $L_{\omega_1\omega}$ but he would look into excellence…

The rest is history: after a few weeks, a first draft of a proof of properties of pseudoexponentiation drawing on a version of excellence and quasiminimality in $L_{\omega_1\omega}$ was circulated, and Zilber started using many methods from excellent classes and infinitary logic. The richness of this approach has provided many interesting connections; we explore some of them in our paper.

A word of thanks from the second author. Once again the second author turns to the first person singular.

I would like to thank Boris Zilber, at a very personal level, for a life-changing conversation we had in 2007 *in Utrecht, during a meeting organized by Juliette Kennedy, on connections between mathematics, philosophy and art. One evening, after dinner, Boris said "let's go for a walk and speak a bit about mathematics." In the cold night along the canals, he described, for about an hour, some of what he had been doing — I kept asking and asking questions. At some point, on a bridge, he turned to me and said, "But you, on what have you been working?" I tried to gather my thoughts on the spot while walking, and started describing a project we had back then with Berenstein and Hyttinen* [11] *of understanding independence notions in continuous logic, trying to extend the work of Chatzidakis and Hrushovski to the continuous case and encountering difficulties. Boris asked me to describe briefly continuous model theory and continuous abstract elementary classes. At some point, he said I obviously had tools for dealing with model theoretical approaches to quantum mechanics. I asked how so. He said, "look at Gelfand triples ... ". I returned to Helsinki, where I was spending a sabbatical, and Boris's remarks made a deep change in my own approach to model theory, in the possibilities I started slowly unfolding. I am deeply grateful for that momentous conversation, and for all the lines of work arising from that evening*!

2. Model theory in mathematics

We first deal with some variations in model theoretic and geometric terminology.

2.1. ***Model theoretic background.*** Mathematical logic makes a central distinction between a vocabulary and a collection of sentences in a logic. For this reason, we use 'language' only for the second and reserve 'vocabulary' for what is sometimes called similarity type.

Definition 2.1.1 (vocabulary and structure).

1. A vocabulary τ is a collection of constant, relation, and function symbols (with finitely many arguments).
2. A τ-structure is a set in which each τ-symbol is interpreted, e.g., an n-ary relation symbol as an n-ary relation.

Definition 2.1.2. *Full formalization* involves the following components.

1. A *vocabulary* with associated notion of structure as in Definition 2.1.1.
2. A *logic* $\mathcal{L}$ has:
   a A class $\mathcal{L}(\tau)$ of 'well formed' *formulas*.
   b A notion of '*truth* of a formula' from the class $\mathcal{L}(\tau)$ in a τ-structure, usually denoted $\mathfrak{A} \models \varphi$.
   c A notion of a 'formal *deduction*' for this logic.
3. *Axioms*: Specific sentences of the logic that specify the basic properties of the situation in question.

Example 2.1.3 (three important logics).

1. The *first order language* $\mathcal{L}_{\omega,\omega}(\tau)$ associated with τ is the least set of formulas containing the atomic τ -formulas and closed under *finite* Boolean operations and quantification over finitely many individuals.
2. The $\mathcal{L}_{\omega_1,\omega}(\tau)$ *language* associated with τ is the least set of formulas containing the atomic τ -formulas and closed under *countable* Boolean operations and quantification over finitely many individuals.
3. The *second order language* associated with τ, denoted $\mathcal{L}^2(\tau)$, is the least set of formulas extending $\mathcal{L}_{\omega,\omega}(\tau)$ by allowing quantification over sets and relations. $\mathcal{L}^2(\{=\})$ is symbiotic ('morally equivalent', roughly speaking) with set theory.

Morley rank (corresponding to the Krull/Weil dimension in the particular case of fields) was introduced in [36] to study theories categorical in uncountable power. Section 5 explores the role of finite Morley rank groups in studying covers. Three good sources for the more advanced model theory used here are [33; 41; 49].

2.2. ***Various viewpoints.*** We now discuss two quite different uses of the three words *automorphism*, *model* and *definable*, coming from areas of mathematics relevant to this paper. (The difference in use depending on the area of mathematics has been at times a source of confusion.)

Remark 2.2.1 (automorphism: two notions).

In model theory: An *automorphism* of a τ-structure $\mathcal{A}$ is a permutation of its universe A that preserves (in both directions) each relation or function symbol for τ. For instance, the automorphisms of a geometry (when given in terms of lines and points together with an incidence relation) are the *collineations*.

In algebraic geometry: An *automorphism* of a variety is an invertible morphism.[2]

[2]This begs the question of defining morphism. A good approximation is "definable map". In algebraic geometry a morphism is a constructible (generically quasirational) bijection; cf. [42, p. 79, Section 4.4]. Biregular and birational are more specific syntactic restrictions on an isomorphism.

Remark 2.2.2 (model: two notions).

In model theory: The word *model* also sees different uses depending on the area. In logic, a model is sometimes just a τ-structure but often signifies that the structure satisfies a theory (as in "$(\mathbb{C}, +, \cdot, 0, 1)$ is a model of the theory ACF_0"). *Minimal model* might mean 'no proper elementary submodel' or, very differently, 'every definable subset is finite or cofinite'.

In algebraic geometry: A *model* is a specific biregularity class within a birational equivalence class. In Weil/Zariski style, a variety is determined by a coordinate ring, but only up to isomorphism of this coordinate ring. A 'model' of the variety might be a specific affine variety with that coordinate ring, but any biregularly isomorphic variety would also be a model.

Thus, unlike model theory, algebraic geometry does not identify 'models' up to isomorphism. Rather, it looks for a specific 'canonical representation' among 'isomorphic solution sets'. A *minimal model* is a smooth variety X with function field K such that if Y is another smooth variety with function field K and $f : X \mapsto Y$ is birational, then f is an isomorphism.

Remark 2.2.3 (definable/defined: two notions).

In model theory: A subset X of a model M^n is *defined* over a set A if there is a formula $\phi(\boldsymbol{x}, \boldsymbol{a})$ with solution set X.

In usual mathematics: the word 'defined' is often short for 'well-defined' saying that the value of a function defined on a quotient space does not depend on the choice of a representative.

In model theory, we add the adjective 'definable' when there is a formula of the language that captures the notion. Thus, the algebraic geometric 'automorphism' becomes 'definable bijection'. It is worth noting that many important automorphisms in algebraic geometry do not necessarily preserve structure.

Remark 2.2.4 (Why infinitary logic?). A natural question at this point is: Why is axiomatizability in $L_{\omega_1,\omega}$ relevant to geometric questions? The answer to this question is not univocal, and strongly reflects different historical issues arising in different areas of mathematics. We discuss four responses, two from ordinary mathematics, two from logic.

1. **In ordinary mathematics:**
   (a) The constraints of expressibility offered by a particular logic force a detailed analysis of the hypotheses of a result. This analysis in similar earlier cases has led to, for example, the Zilber–Pink conjecture and the Conjecture on the intersection of tori (see, e.g., [13]).

(b) Of course, each of the 'canonical structures' is explicitly definable in set theory. But this definition in most cases is useless for studying the object. Useful succinct second order axioms are available for the real and complex numbers but are only partially known for universal covers. First order logic is stymied *a priori* by the intractability of arithmetic. Thus, categoricity in infinitary logic is essential for *giving an 'algebraic' account of an 'analytic object'*. This use of model theory can be seen as part of the larger scale *GAGA* mathematical program of bridging analytical concepts and algebraic ones.

2. **In logic (in particular, in model theory):**
   (a) A natural question is: are there important mathematical notions expressible in infinitary logic which are not expressible in first order? The study of complex exponentiation yielded a superb initial example: the categoricity of the covering map of $\mathbb{C}^*$ in [6].
   (b) This raises the question of what are the *new* axioms in this paper that require an infinitary description. The infinite dimension axioms are well known and the switch from 'standard kernel' to 'standard fiber over z' (i.e., $q^{-1}(z)$) is unremarkable. It seems the finite index conditions (Section 4.4) are not first order expressible.

3. Categoricity, quasiminimality and excellence

We give a quick sketch of notions around categoricity[3] and the history of their logical development.

Definition 3.0.1 (categoricity).

1. A *theory* T in a logic $\mathcal{L}$ is a collection of $\mathcal{L}$-sentences in a vocabulary τ.
2. T is *categorical in cardinality* κ (κ-categorical) if all models M of T with $|M| = \kappa$ are isomorphic.

Although certain canonical mathematical structures are fruitfully axiomatized in second order logic, rather than second order categoricity, we usually consider these characterizations as defining these structures *in set theory.* Such definitions are exactly what it means to be a structure. Second order categoricity *per se* gives no useful mathematical information. In contrast, κ-categoricity in first order logic or in $L_{\omega_1,\omega}$ provides very significant (combinatorial geometric) information; it assigns a dimension to each model.

[3] More specifically, when in model theory we use the word *categoricity,* we mean categoricity in a specific cardinality or 'in power'. See a thorough discussion of categoricity in various logics in [4, §3.1] and an exposition of the philosophical import of the notion in [14].

3.1. ***The classical categoricity theorems.*** The following results survey the spectrum of cardinals in which certain types of theory can be categorical. These theorems are of the form *if a theory (or a sentence) is categorical in* <u>*some high enough*</u> *cardinal(s), then it must be categorical on* <u>*a tail of*</u> *cardinals.*

Theorem 3.1.1 (Morley's categoricity theorem [36]). *A countable first order theory is categorical in one uncountable cardinal if and only if it is categorical in all uncountable cardinals.*

Theorem 3.1.2 (Shelah's categoricity under the weak continuum hypothesis below $\aleph_\omega$ [45; 46]). *Assuming* $2^{\aleph_n} < 2^{\aleph_{n+1}}$ *a* ***sentence*** *in* $L_{\omega_1,\omega}$ *that is categorical in* $\aleph_n$ *(for every* $n < \omega$*) is categorical in all uncountable cardinals.*

Theorem 3.1.3 (Shelah's categoricity theorem for excellent sentences [45; 46]). *An* ***excellent sentence*** *in* $L_{\omega_1,\omega}$ *is categorical in one uncountable cardinal if and only if it is categorical in all uncountable cardinals.*

Theorem 3.1.4 (Zilber's categoricity for quasiminimal excellent classes). *A quasiminimal excellent class is categorical in all uncountable cardinals* [54].

3.2. ***Pregeometries (matroids) and quasiminimality.*** The presence of quasiminimal pregeometries provides an extremely fruitful and natural control of models in a class (and of their interactions).

Definition 3.2.1 (combinatorial geometry). A *closure system* is a set G together with a 'closure' relation on subsets of G,

$$\mathrm{cl} : \mathcal{P}(G) \to \mathcal{P}(G),$$

satisfying the following axioms.

A1. $\mathrm{cl}(X) = \cup\{\mathrm{cl}(X') : X' \subseteq_{\mathrm{fin}} X\}$.

A2. $X \subseteq \mathrm{cl}(X)$.

A3. $\mathrm{cl}(\mathrm{cl}(X)) = \mathrm{cl}(X)$.

(G, cl) is a *pregeometry* if, in addition, we have:

A4. If $a \in \mathrm{cl}(Xb)$ and $a \notin \mathrm{cl}(X)$, then $b \in \mathrm{cl}(Xa)$.

If points are closed ($\mathrm{cl}(\{a\}) = \{a\}$, for each a) the structure is called a *geometry*.

Pregeometries are virtually the same mathematical objects as matroids.

Definition 3.2.2.
1. A subset D of a τ-structure M is *first order-definable* in M if there is $\boldsymbol{a} \in M$ and an $\mathcal{L}_{\omega,\omega}(\tau)$-formula $\varphi(x, \boldsymbol{y})$ such that $D = \{m \in M : M \models \varphi(m, \boldsymbol{a})\}$. If $\boldsymbol{a} \in A \subseteq M$, D is definable with parameters from A.
2. $\mathrm{acl}_M(A)$, the *algebraic closure* of A in M, is $\{m \in M : \phi(m, \bar{a}), \bar{a} \in A\}$, where $\phi(x, \bar{a})$ has only finitely many solutions in M.

3. $\mathrm{dcl}_M(A)$, the *definable closure* of A in M, is defined as was the algebraic closure, but replacing 'finitely many' by 'one'.
4. An infinite definable subset D (or its defining formula $\varphi(x)$) is *strongly minimal* if every definable subset of D in every elementary extension of M is finite or cofinite.
5. A theory is *strongly minimal* if the formula $x = x$ is strong minimal.

The notion of type is a crucial tool in model theory.

Definition 3.2.3. 1. The *first order type* of a over B (in M), denoted $\mathrm{tp}_M(a/B)$, is the set of $\mathcal{L}_{\omega,\omega}$-formulas with parameters from B that are satisfied in M (for $a, B \subseteq M$).

2. The *quantifier-free type* of a over B (in M), denoted $\mathrm{tp}_{\mathrm{qf}}(a/B\colon M)$, is the set of *quantifier-free* first order formulas $\varphi(x, \boldsymbol{b})$ such that $M \models \varphi(a, \boldsymbol{b})$ (as before, $\boldsymbol{b}$ ranges over tuples of B).

In most contexts, when we just say 'the type of a over B,' we mean the first order type. Note also that if a property is defined without parameters in M, then it is uniformly defined in all models of $\mathrm{Th}(M)$ (the *theory* of M, i.e., the set of all τ sentences that are true in M).

Here are three fundamental observations on strongly minimal sets.

- A strongly minimal set admits a combinatorial geometry when the closure is taken as acl (Definition 3.2.2).
- There is a unique type of elements in a strongly minimal set that are not algebraic. This is called the *generic type* for D.
- In many important examples (e.g., DCF_0), the structure of the model is controlled by its strongly minimal sets.

Shelah's abstract notion of independence (for some first order theories, crystallized as *nonforking*) weakens the notion of combinatorial geometry by dropping **A3**; in some desirable cases this property is recovered on the points realizing a *regular type* and in even better cases the dimensions of the regular types determine the isomorphism type of the model. However, *a priori*, the existence of a global dimension is unusual.

We now look at the generalization of strong minimality, introduced by Zilber, that is central in the connections between model theory and algebraic geometry described in this paper.

Definition 3.2.4 (quasiminimal structure). A structure M is *quasiminimal* if every first order ($L_{\omega_1,\omega}$) definable subset of M is countable or cocountable. Algebraic closure is generalized by saying $b \in \mathrm{acl}'(X)$ if there is a first order formula with *countably many* solutions over X which is satisfied by b.

Definition 3.2.5 (quasiminimal excellent geometry). Let $\boldsymbol{K}$ be a class of L-structures such that $M \in \boldsymbol{K}$ admits a closure relation cl_M mapping $X \subseteq M$ to $\mathrm{cl}_M(X) \subseteq M$ that satisfies the following properties.

1. ***Basic conditions***
   (a) Each cl_M defines a pregeometry on M.
   (b) For each $X \subseteq M$, $\mathrm{cl}_M(X) \in \boldsymbol{K}$.
   (c) (the countable closure property, or ccp): If $|X| \leq \aleph_0$ then $|\mathrm{cl}(X)| \leq \aleph_0$.
2. ***Homogeneity***
   (a) A class $\boldsymbol{K}$ of models has $\aleph_0$*-homogeneity over* $\varnothing$ (Definition 3.2.5) if the models of $\boldsymbol{K}$ are pairwise qf-back and forth equivalent (Definition 4.3.7).
   (b) A class $\boldsymbol{K}$ of models has $\aleph_0$*-homogeneity over models* if for any $G \in \boldsymbol{K}$ with G empty or a countable member of $\boldsymbol{K}$, any H, H' with $G \leq H$, $G \leq H'$, H is qf-back and forth equivalent with H' over G.
3. $\boldsymbol{K}$ is an *almost quasiminimal excellent geometry* if the universe of any model $H \in \boldsymbol{K}$ is in $\mathrm{cl}(X)$ for any maximal cl-independent set $X \subseteq H$.
4. We call a class which satisfies these conditions an *almost quasiminimal excellent geometry* [8].

An almost quasiminimal excellent geometry with strong submodel taken as $A \leq M$, if $\mathrm{acl}_M(A) = A$, gives an *abstract elementary class* (AEC)[4]. But the distinct notion of a quasiminimal AEC (defined in terms of $\leq$ rather than any axioms) is due to Vasey [50].

To obtain that the class is complete for $L_{\omega_1,\omega}$, [8; 30] add the requirement of $\aleph_0$-categoricity.

Remark 3.2.6. This definition differs only superficially from those in, e.g., [30], where the connections with the combinatorial geometry was emphasized by distinguishing the treatment of elements depending on whether they were in $\mathrm{cl}(H)$. However, [8] required a quasiminimal structure to have a unique generic type. This requirement fails in the two-sorted treatment we deal with here; there may be acl-bases in each sort. So we replace quasiminimality with *almost quasiminimality* (less explicit in [9]) and we thus restore Zilber's first intuition (Definition 3.2.4) that quasiminimality means that all definable sets are countable or cocountable.

Remark 3.2.7 (excellence). From Zilber's introduction [54] of the notion, it has been known that the axioms 3.2.5 imply $\aleph_1$-categoricity. See the exposition in [3]. But, without further 'excellence' hypotheses, it was unknown whether the class had larger models. Two formulations of excellence are: (1) [45; 46] n-amalgamation of independent systems of models, for all $n < \omega$; (2) [30] a local condition on the

[4]See [23] for the early history of the model theory of AECs.

properties of a 'crown'. Either implies the existence of arbitrarily large models for theories in $L_{\omega_1,\omega}$. As we discuss in Section 5.2, influenced by work Hart and Shelah on first order classification theory, the next result (here modified by 'almost') clarified the relationship.

Crucial Fact 3.2.8 (Bays, Hart, Hyttinen, Kesälä, Kirby). *Every almost-quasi-minimal class* (Definition 3.2.5) *is excellent* (in the sense of Remark 3.2.7). *Thus, it is categorical in all uncountable cardinalities.*

4. Modular and Shimura curves

We begin with an astronaut's view of the j-function and then turn to the model theoretic treatment of some generalizations.

4.1. *The great confluence.* The general form (over a field of characteristic 0) of an elliptic curve is

$$y^2 = x^3 + ax + b.$$

At least since Diophantus (3rd century AD), the search for integer solutions for such equations has been a central question. The cataloguing of such equations was a major achievement of the 19th century. One key step toward this classification is to generalize the original problem and look first for complex solutions. The solution set of an elliptic curve is then a smooth, projective, algebraic curve of genus one. It can be thought of as a 'classical torus' $\mathbb{T}_\tau := \mathbb{C}/\Lambda_\tau$, where $\tau \in \mathbb{C}$ and Λ_τ is the lattice in $\mathbb{C}$ (the subgroup of $(\mathbb{C}, +)$ generated by $\langle 1, \tau\rangle$.

Klein studied modular and automorphic functions, which provide surprising and deep links between geometry, complex analysis and number theory. The most famous example is the j-function, analytic on $\mathbb{H} = \{z : \mathrm{Im}(z) > 0\}$, the upper half plane, and maps onto $\mathbb{C}$ and meromorphic with some poles on the real axis and the following remarkable properties.

Theorem 4.1.1 (classification of tori by the j-function). *The following conditions are equivalent*:

1. *There exists* $s = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{Z})$ *such that* $s(\tau) = (a\tau + b)/(c\tau + d) = \tau'$.
2. $\mathbb{T}_\tau \approx \mathbb{T}_{\tau'}$ (in the algebraic geometry sense of Remark 2.2.1.)
3. $j(\tau) = j(\tau')$.

This astonishing classical fact paves the way toward modern day classifications. It provides equivalences between analytic and number-theoretic notions. Strikingly, j is defined as a rational function of two analytic functions g_2 and g_3 (each of them coding so-called 'modularity' properties):

$$j(\tau) = 12^3 \cdot \frac{g_2(\tau)^3}{g_2(\tau)^3 - 27g_3(\tau)^3}.$$

But where does the word ‘elliptic’ come from? A meromorphic function is called an *elliptic function* if it is doubly periodic: there are two $\mathbb{R}$-linear independent complex numbers ω_1 and ω_2 such that $\forall z \in \mathbb{C}$, $f(z+\omega_1) = f(z)$ and $f(z+\omega_2) = f(z)$. Abel discovered such doubly periodic functions arose from the solutions of elliptic integrals — originally defined to find the arc length of an ellipse. Weierstrass used the symbol $\wp$ to denote a family of functions $\wp(z, \Lambda_\tau)$ where the defining double sum runs over the elements of the lattice Λ_τ, generated by 1 and τ. The crucial property of the function is that every meromorphic function that is periodic on Λ_τ is a rational combination of $\wp(z, \Lambda_\tau)$ and $\wp'(z, \Lambda_\tau)$. This field of functions is precisely Abel’s field of elliptic functions.

Klein’s discovery of the j function unified the results of Weierstrass. In his famous investigation of the psychology of mathematical investigation, Hadamard devotes several pages to Poincaré’s generalization of the j-function to the family of functions derived from Fuchsian group actions. The crucial phrase for us is *‘the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry’* [24, p. 33].

This completes a very quick summary of the 19th century predecessors of the theory of moduli spaces, developed in the next section. This study involves complex analysis, actions by a discrete group, number theory, and non-Euclidean geometry. The crucial model theoretic step is to formalize in a vocabulary for two-sorted structures of the form

$$\mathfrak{A} = \big\langle \langle H; \{g_i\}_{i\in\mathbb{N}}\rangle, \langle F, +, \cdot, 0, 1\rangle, j : H \to F \big\rangle,$$

where $\langle F, +, \cdot, 0, 1\rangle$ is an algebraically closed field of characteristic 0, $\langle H; \{g_i\}_{i<\omega}\rangle$ is a set together with countably many unary function symbols, and $j\colon H \to F$.

In the next section we provide some of the mathematical background for a formal analysis of these two-sorted structures.

4.2. ***Moduli spaces.*** Moduli spaces in geometry are parametrized collections of objects, together with equivalences that allow us to see when two objects are in some sense ‘the same’, and with families that articulate the variation between the objects in the collection. Paraphrasing the important survey [10], ‘moduli spaces are a *geometric* solution to a geometric classification problem.’ They parametrize collections of geometric *objects*, they define *equivalences* to say when two objects are the ‘same’, and establish *families* that determine how we allow our objects to vary or modulate.

In model theory, the notion of a uniform family of definable sets has been thoroughly studied. Such a family is given by a formula of the form $\phi(\boldsymbol{x}, \boldsymbol{y})$. Each set in the family is the solution set of $\phi(\boldsymbol{a}, \boldsymbol{y})$ (for some $\boldsymbol{a}$), and the set $\{\boldsymbol{a} : (\exists \boldsymbol{y})\phi(\boldsymbol{a}, \boldsymbol{y})\}$ is an indexing set of the family. In the algebraic geometry setting, one can require that the $\boldsymbol{x}$ fall into a variety V and the $\boldsymbol{y}$ into a variety $W_{\boldsymbol{a}}$. V is a step toward the notion of a *moduli space*.

Except in Section 5, we consider moduli spaces arising from a pair (G, X) consisting of a group G acting on a space X. The algebraic varieties we study arise as quotients $\Gamma \setminus X$ (for Γ a subgroup of G; see Definition 4.2.2). A modular curve arises as a connected component of quotient of $\mathbb{H}$ by congruence subgroups (Definition 4.2.9) of $\mathrm{GL}_2(\mathbb{R})$. Shimura generalized the topic to groups acting on wider classes of domains. Shimura curves are rather more complicated yet generally share similar categoricity properties. Shimura varieties of higher dimension raise many new issues that we sketch in Section 6. In this section, we consider only covers of modular curves by $\mathbb{H}$.

Here, $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ refers, as in the rest of this paper, to the upper half complex plane, also called the hyperbolic plane when endowed with a metric and topology that make it hyperbolic rather than Euclidean. See [35] for a detailed description. In all our examples, the function p maps the hyperbolic plane into a complex variety. We consider the action of $\mathrm{PSL}_2(\mathbb{R})$ on $\mathbb{H}$ as fractional linear transformations: for $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{Z})$ and $\tau \in \mathbb{H}$, $A(\tau) = (a\tau + b)/(c\tau + d)$.

The group of bijections (isometries, $\mathrm{isom}(\mathbb{H})$) that preserve the hyperbolic metric of $\mathbb{H}$ is generated by $\mathrm{PSL}_2(\mathbb{R})$ and the map $z \mapsto -\bar{z}$; $\mathrm{PSL}_2(\mathbb{R})$ consists precisely of all those isometries that preserve orientation (e.g., [28]). After outlining here the classical theory of such actions and moduli spaces, in Section 4.3 we describe a model theoretic approach.

Definition 4.2.1 (Fuchsian group).

1. A subgroup $G \leq \mathrm{isom}(\mathbb{H}) \approx \mathrm{PSL}_2(\mathbb{R})$ is *discrete* if it is discrete in the induced topology.
2. A *Fuchsian group* is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$.

The most important example of a Fuchsian group is $\mathrm{PSL}_2(\mathbb{Z})$. Underlying this entire study and almost one and a half centuries of interactions between number theory and complex analysis is the remarkable fact that the quotient of $\mathbb{H}$ by certain discrete subgroups has the structure of a Riemann surface [35, §1.8] and even an algebraic variety which, in important cases, is a moduli space [34].

Definition 4.2.2 (quotient of $\mathbb{H}$ by a group). If a group G acts on a set X, $G \setminus X$ has universe the collection of G-orbits of the action. π is the canonical map taking x to its orbit Gx. The prototypical example corresponds to $X = \mathbb{H}$.

Definition 4.2.3. The quotients $V = S(\mathbb{C})$ of $\mathbb{H}$ by a discrete group Γ that we consider are examples of *moduli spaces*. $V = \bigcup_{a \in C} V_a$ is the image of a map p from $\mathbb{H}$ that acts as a *uniformizer* for a family of varieties V_a. Namely, for each

$a, b \in \mathbb{H}$, we have $V_a \cong V_b$ if and only if, for some $\gamma \in \Gamma$,

$$\gamma(a) = b \iff p(a) = p(b).$$

We explored in Section 4.1 the ur-example of a moduli space, elliptic curves as uniformized by the j-function. The next definition relies on the fact that, while elements of $\mathrm{PSL}_2(\mathbb{R})$ fix $\mathbb{H}$ setwise, they also act on all of $\mathbb{C}$.

Definition 4.2.4 (cusp). Let Γ be a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$.

1. We say $c \in \mathbb{R} \cup \{\infty\}$ is a *cusp* of Γ if c is the unique fixed point of some $\gamma \in \Gamma$.
2. P_Γ is the set of cusps of Γ and $\mathbb{H}^* = \mathbb{H}^*_\Gamma = \mathbb{H} \cup P_\Gamma$.

We relate some standard facts (see [25, p. 15]). The first relies on the fact that while some of the quotients we study are not compact, they can be compactified by adding finitely many cusps from $\mathbb{R} \cup \{\infty\}$.

Fact 4.2.5. For any discrete subgroup $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$, the quotient $\Gamma \setminus \mathbb{H}^*_\Gamma$ is a compact Hausdorff space that can be given the structure of a Riemann surface. Therefore if Γ' is of finite index in Γ, the quotient $\Gamma' \setminus \mathbb{H}^*_\Gamma$ is a compact Riemann surface, and is therefore algebraic by the Riemann existence theorem. $\mathbb{H}^*_\Gamma$ is the compactification of the quasiprojective algebraic variety (so first order definable) $\mathbb{H}_\Gamma$.

For the purposes of this paper, since the quasiprojective variety $\mathbb{H}_\Gamma = \Gamma\backslash\mathbb{H}$ determines the (classical) algebraic variety (set of solutions of a system of polynomial equations) $\mathbb{H}^*_\Gamma$, we work hereafter with $\mathbb{H}_\Gamma$. This is natural from a model-theoretic standpoint since (in this situation) there are only finitely many cusps and so the sets differ by only finitely many points.

Notation 4.2.6 *fixes the group G for the rest of Section 4.* Setting the determinant as 1 and modding out the center guarantees the group action preserves both distance and orientation.

Notation 4.2.6. Let $G = \mathrm{GL}_2^{ad}(\mathbb{Q})^+ =_{\mathrm{def}} \mathrm{PSL}_2(\mathbb{Q})/Z(\mathrm{PSL}_2(\mathbb{Q})) \approx \mathrm{PSL}_2(\mathbb{Q})$ modulo its center. Γ varies over subgroups of G

We now distinguish two kinds of points in $\mathbb{H}$: 'special' points and 'Hodge-generic' points. The equivalence of the following definition with the usual notion [15, Definition 2.2] for Shimura varieties is in [15, Theorem 2.3].

Definition 4.2.7 (special points). Fix $\langle \mathbb{H}, S(\mathbb{C}), p \rangle$ with $S(\mathbb{C})$ biholomophic to $\Gamma \setminus \mathbb{H}$. A point $x \in \mathbb{H}$ is *special* if there is a $g \in G$ whose unique fixed point is x.

We omit the definition of a Hodge generic point arising in algebra, as it does not enter our discussion; we use only the equivalent characterization given in part 1 of Fact 4.2.8 and the dichotomy in part 2. (The equivalence is part of Proposition 2.5 in [15] and the dichotomy is noted just after that proposition.) It is worth mentioning that for a point the fact of being 'special' or 'Hodge generic' does not depend on the choice of the group Γ; furthermore, these two notions are preserved by the action of $G = GL_2^{\mathrm{ad}}(\mathbb{Q})^+$.

Fact 4.2.8 (special and Hodge generic points [15, Proposition 2.5]).

1. If x is Hodge generic the only $g \in G$ that fixes x is the identity.
2. Every point in $\mathbb{H}$ is either Hodge generic or special.

Although we are studying the categoricity of the universal cover of a specific modular curve (e.g., the image of the j-function, $\Gamma \setminus \mathbb{H}$), other modular curves naturally arise in the analysis. The study of families of such curves is expounded in [48, Sections 6 and 7]. A key tool to give a uniform treatment to a family is the existence of a common commensurator of the generating Fuchsian groups. In fact, the members of the family are interalgebraic and the entire family (indexed by the Γ_N) is studied in [17].

Definition 4.2.9. 1. The groups Γ_N (N a fixed integer) are given by

$$\Gamma_N = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma : b \equiv c \equiv 0,\ a \equiv d \equiv 1 \bmod N \right\}.$$

Note that each Γ_N has finite index in Γ and if $N|M$ then $\Gamma_M \subseteq \Gamma_N$.

2. Two subgroups Γ and Γ' of a group H are said to be *commensurable* if $\Gamma \cap \Gamma'$ is of finite index in both of them.
3. A *congruence subgroup* is a subgroup Γ' of Γ such that some Γ_N is a finite index subgroup of Γ'.
4. The *commensurator* $\mathrm{comm}(\Gamma)$ of a subgroup Γ of $\mathrm{PSL}_2(\mathbb{R})$ is

$$\{\delta \in \mathrm{PSL}_2(\mathbb{R}) : \delta\Gamma\delta^{-1} \text{ is commensurable with } \Gamma\}.$$

We rely on the following standard fact.

Lemma 4.2.10. *The group $G = \mathrm{GL}_2^{ad}(\mathbb{Q})^+$ (Notation 4.2.6) is the commensurator of any congruence subgroup Γ of $\mathrm{SL}_2(Z)$.*

Because the functions $g \in G$ are in the formal vocabulary, we employ congruence subgroups $\Gamma_{\boldsymbol{g}}$ from Notation 4.2.11 rather than the Γ_N. The $Z_{\boldsymbol{g}}$ defined in Notation 4.2.11 play a central role both in the quantifier elimination and via an inverse limit in Section 4.4.

Notation 4.2.11. Let G be fixed as in Notation 4.2.6, and recall that each of the congruence subgroups of $\mathrm{PSL}_2(\mathbb{Z})$ act on $\mathbb{H}$. For any finite sequence $\boldsymbol{g} = \langle e, g_2, \ldots, g_n \rangle$ from G (by convention, $g_1 = e$), introduce the following objects, which are well-defined by our choice of p and Γ.

1. $\Gamma_{\boldsymbol{g}} = \Gamma \cap g_2^{-1}\Gamma g_2 \cap \cdots \cap g_n^{-1}\Gamma g_n$.

2. Let $p : \mathbb{H} \to S(\mathbb{C})$.
   (a) $Z_{\boldsymbol{g}}$ is defined as $\{(p(x), p(g_2 x), \ldots, p(g_n x)) \in S(\mathbb{C})^n : x \in \mathbb{H}\}$.
   (b) $p_{\boldsymbol{g}} : \mathbb{H} \to Z_{\boldsymbol{g}} \subseteq S(\mathbb{C})^n$ is defined by

$$x \mapsto p(\boldsymbol{g}x) = \langle p(x), p(g_2(x)), \ldots, p(g_n(x))\rangle.$$

   (c) $[\phi_{\boldsymbol{g}}] : \mathbb{H}_{\boldsymbol{g}} \to Z_{\boldsymbol{g}}$ is defined by $[\phi_{\boldsymbol{g}}]x_{\Gamma_{\boldsymbol{g}}} = p_{\boldsymbol{g}}(x)$; by Lemma 4.2.12, it is onto.
3. $\mathbb{H}_{\boldsymbol{g}} = \Gamma_{\boldsymbol{g}} \setminus \mathbb{H}$.

The following lemma is central to Section 4.4.2. Its proof uses Shimura theory very heavily.

Lemma 4.2.12 [19, 3.22]. *The map* $[\phi_{\boldsymbol{g}}]$ *is bijective on the Hodge generic points and the image* $Z_{\boldsymbol{g}}$ *is a* ***variety*** *contained in* $\mathcal{S}^n(\mathbb{C})$, $n = \lg(\boldsymbol{g})$. *Moreover, for all* $\boldsymbol{g}$, $Z_{\boldsymbol{g}}$ *is defined over the maximal abelian extension* L *of the field of definition*, E, *of* S.

Remark 4.2.13. From the model theoretic standpoint, it makes no sense to say the $[\phi_{\boldsymbol{g}}]$ are definable since their domains $\mathbb{H}_{\boldsymbol{g}}$ are not. While the maps $[\phi_{\boldsymbol{g}}]$ are bijective on Hodge generic points, they may identify special points.

4.3. *First order completeness for modular and Shimura curves.* We now lay out the vocabulary and first order theory for studying modular curves. The mathematical input is a Fuchsian group Γ acting on hyperbolic space $\mathbb{H}$ and the image curve $S(\mathbb{C}) = \Gamma \setminus \mathbb{H}^*_\Gamma$ (Definition 4.2.4) with a standard model $\boldsymbol{p} = \langle \mathbb{H}, S, p\rangle$. The structure of a discrete group is unwieldy from a traditional model theoretic standpoint because its first order theory is unstable and undecidable. Just as modules are usually studied in model theory by adding unary function symbols f_r for the elements of the ring, in order to represent the action of G on $\mathbb{H}$, we add symbols f_g for $g \in G$ as unary functions that act on $\mathbb{H}$. We thus use a two-sorted presentation of our structures: a sort for the domain, a sort for the target, and a map p connecting them.

Remark 4.3.1 (sorts). A two-sorted structure interprets two sort symbols and additional relation and function symbols with the understanding that each such relation/function either is restricted to one of the predicates or explicitly connects them.

Notation 4.3.2 (the formal vocabulary τ). The two-sorted vocabulary τ consists of the sorts (unary predicate symbols) D (the covering sort), S the target sort, and a function q mapping D onto the sort S.

We write τ_G for the vocabulary of the first sort with $G = G^{ad}(\mathbb{Q}^+)$. The second $\tau_F = \mathcal{R}$ where $\mathcal{R}$ is the set of formulas in $\{+, -, 0, 1, \times\}$ specified in Notation 4.3.3. τ is $\tau_G \cup \tau_F \cup \{p\}$. There are *constant symbols* for each element of the field $E^{ab}(\Sigma)$

defined in Notation 4.3.3. We use f_g to name the functions acting on D, but often write the shorter $g(x)$ or gx instead of $f_g(x)$.

The following notation is essential to understand the Axioms 4.3.5. Note in the prototype q is replaced by the known covering map p.

Notation 4.3.3. The *standard model* for a *modular curve* determined by a Fuchsian group $\Gamma \subseteq G = G^{ad}(\mathbb{Q}^+)$ will consist of a τ-structure $\boldsymbol{p} = \langle \mathbb{H}, S, p\rangle$ with the domain $\mathbb{H}$, the variety $S(F)$ over the algebraically closed field F defined by $\Gamma \setminus \mathbb{H}$, and $\mathcal{R}$ the set of all Zariski closed relations on $S(F)^n$ (for all n) with constants from a field $E^{ab}(\Sigma)$ that are true in F. E^{ab} is the maximal abelian extension of the defining (reflex) field E of S. $E^{ab}(\Sigma)$ is the extension of E^{ab} (F_0 in [19, §4, p. 17]) obtained by adding the coordinates of the ($\leq \aleph_0$) special points, and closing to a field.

Notation 4.3.4. For a structure $\boldsymbol{p}$, we write $\mathrm{Th}(\boldsymbol{p})$ for the complete first order theory of all sentences true in $\boldsymbol{p}$ and $T(\boldsymbol{p})$ for the specified set of axioms true of $\boldsymbol{p}$. Clearly, $T(\boldsymbol{p}) \subseteq \mathrm{Th}(\boldsymbol{p})$.

We must distinguish $\mathrm{Th}(\boldsymbol{p})$ from its subset $T(\boldsymbol{p})$ until we prove $T(\boldsymbol{p})$ is a complete axiomatization of $\mathrm{Th}(\boldsymbol{p})$.

Definition 4.3.5 (first order axioms). $T(\boldsymbol{p})$ is the following collection of first order sentences that are to hold in a structure $\langle D, S(F), q\rangle$.

1. Each sentence in $\mathrm{Th}(\langle \mathbb{H}, \{f_g : g \in G\rangle)$. These include 'special point axioms' SP_g: For each $g \in G$ that fixes a unique point in D,
$$\forall x, y \in D[(g(x) = x \wedge g(y) = y) \Rightarrow x = y].$$
2. $\mathrm{Th}(S(\mathbb{C}), \mathcal{R})$ ($\mathcal{R}$ from Notation 4.3.2).
3. The covering map; for each $\boldsymbol{g} \in G^m$ and all $m < \omega$:
   (a) $\mathrm{Mod}^1_{\boldsymbol{g}}$: $\quad \forall x \in D\ (q(g_1(x), \ldots q(g_m(x)) \in Z_{\boldsymbol{g}})$.
   (b) $\mathrm{Mod}^2_{\boldsymbol{g}}$: $\quad \forall z \in Z_{\boldsymbol{g}}\ \exists x \in D\ (q(g_1(x)), \ldots q(g_m(x)) = z)$.
   (c) $\mathbf{MOD} = \{\mathrm{Mod}^1_{\bar{g}} \wedge \mathrm{Mod}^2_{\bar{g}} : \bar{g} \in G^m, m < \omega\}$.

Note that **MOD** is a countable collection of first order sentences.

Notation 4.3.6. By the choice of $E^{ab}(\Sigma)$, special points belong to $\mathrm{dcl}(\emptyset)$. Therefore, we can name each one of them by d_g, where $g \in G$ fixes d_g. Any g that fixes a point is in $G - \mathrm{Sl}_2(\mathbb{Z})$ [19, Lemma 3.13]. There will be distinct g_1, g_2 that fix the same point (e.g., if $g_2 = g_1^2$). If so, $T(\boldsymbol{p}) \vdash d_{g_1} = d_{g_2}$ The theory of (D, G) contains the uniqueness axiom (Definition 4.3.5.1) that entails $g(d_g) = d_g$.

The cover sort is a set with unary functions. Both its theory (since the universe is a union of orbits) and that of the field sort (since algebraically closed) are strongly minimal and quantifier eliminable.

Definition 4.3.7. We say two structures M and N are qf-back and forth equivalent if the system I of partial isomorphisms of M and N between isomorphic *finitely generated substructures* satisfies the back and forth condition: For each $f \in I$ and each $m \in M - \operatorname{dom} f$, there exists an $n \in N$ such that $f \cup \{\langle m, n\rangle\} \in I$, and symmetrically, for each $n \in N - \operatorname{im} f$, there exists $m \in M$ such that $f \cup \{\langle m, n\rangle\} \in I$. In this situation $\operatorname{dom} f$ is definably close.

Notation 4.3.8. We write $\boldsymbol{g}(x)$ for $(g_1(x), \dots g_n(x))$, where $\boldsymbol{g}$ has length n and begins with e. And then $\boldsymbol{g}(\boldsymbol{x})$ denotes the sequence of length nm obtained when $\boldsymbol{g}$ is applied to each element of a sequence $\boldsymbol{x} \in (D)^m$. When convenient we write gx or $\boldsymbol{g}x$ for the action, omitting the parentheses.

We now sketch the proof of Theorem 4.3.13 that $T(\boldsymbol{p})$ axiomatizes a complete, quantifier eliminable τ-theory.

Definition 4.3.9 (the back and forth). Fix two models $\boldsymbol{q} = \langle D, S(F), q\rangle$ and $\boldsymbol{q}' = \langle D', S(F'), q'\rangle$ of $T(\boldsymbol{p})$. We define the qf-back-and-forth system I of substructures of $\boldsymbol{q}$ and $\boldsymbol{q}'$ For each $f \in I$, $\operatorname{dom} f$ and $\operatorname{rg}\ f$ are each finitely generated over $E^{ab}(\Sigma)$. A typical member f of the system for $\boldsymbol{q}$ has $\operatorname{dom} f = U = U_D \cup U_S$. Since U is finitely generated, U_D consists of the G-orbits of a finite number of $x \in D$; U_S is $S(L_U)$ where L_U is the field generated by $E^{ab}(\Sigma)$ (since the elements of $E^{ab}(\Sigma)$ elements are named), the coordinates of the $q(x)$ for $x \in U_D$ and finitely many additional points of $F \cap U$. Note that the additional points determine finitely many new field elements since q is constant on each orbit, so the field remains finitely generated. Define a similar subsystem for $\boldsymbol{q}'$, labeling by putting primes on corresponding objects. By Fact 4.2.8 every point of D is either special and so named in the vocabulary (Notation 4.3.6), or Hodge generic. Thus we can ignore the special points in building the back and forth system.

Suppose f is an isomorphism between $U \subseteq \boldsymbol{q}$ and $U' \subseteq \boldsymbol{q}'$. Then f restricts to a G-equivariant (elements in the same orbit have the same image) injection of U_D into $U_{D'}$ and an embedding of $S(L_U)$ into $S(F')$ induced by an embedding σ of L into $S(F')$, that fixes $E^{ab}(\Sigma)$.

The following claim is stated for arbitrary finite sequences $\boldsymbol{g}$, but only singleton x. The type r_d of an infinite sequence (here represented by an infinite tuple of variables $\boldsymbol{v}$) includes the types of $\boldsymbol{g}x$ for any finite $\boldsymbol{g}$.

The main consequence of the following claim is that we may reduce types of points in the domains sort to quantifier-free types of their images in the field sort.

Claim 4.3.10 [15, Proposition 3.3]. *If $d \in D - U_D$ is Hodge generic*:

$$r_d(\boldsymbol{v}) \models \operatorname{tp}_{qf}(d/U),$$

where $r_d(\boldsymbol{v}) = \bigcup_{g\in G} \operatorname{tp}_{qf}(q(\boldsymbol{g}(d))/U) = \operatorname{tp}_{qf}(\langle q(g(d)) : g \in G\rangle/U)$.

Proof. We show that there is a unique quantifier-free type over U of an element of D that restricts to r_d. The consistent nontrivial types in τ_G are (i) $\{x \neq f : f \in U_D\}$ and (ii) $\{x \neq gx\}$ for any nonidentity $g \in G$. The first is captured by $(q(x), q(f)) \notin Z_{e,e}$ for each $f \in U_D$ and the second by $(q(x), q(x)) \notin Z_{e,g}$ if $g \notin \Gamma$ and these are both in $r(\boldsymbol{v})$.

Suppose $\boldsymbol{h} \in S(M)^\omega$ (for a saturated $M \models T(\boldsymbol{p})$ containing U) realizes $r_d(\boldsymbol{v})$ and $\boldsymbol{h}$ with $d' \in D(M)$ satisfy $\boldsymbol{h} = \langle q(g(d')) : g \in G\rangle$. By the previous paragraph $d' \notin U_D$. So d' realizes $\mathrm{tp}_{qf}(d/U)$ as required. □

Notation 4.3.11. For a type $r(v)$ over a set A and an isomorphism f from A to B, $f(r)$ is the set of B-formulas $\phi(v, f(\boldsymbol{a}))$ with $\phi(v, \boldsymbol{a}) \in r$.

Claim 4.3.12 [15, Proposition 3.4]. *Fix $\boldsymbol{g}$. If $x \in U_D$, there is an $x' \in U_{D'}$ such that $q(\boldsymbol{g}(x')) \in S(F')^m$ realizes $f(\mathrm{tp}_{qf}(q(\boldsymbol{g}(x))/L_U))$.*

Proof. We write $Z^{\boldsymbol{q}}_{\boldsymbol{g}}$ for the points in $S(F)$ satisfying (the formula defining) $Z_{\boldsymbol{g}}$. Using Notation 4.3.11, Claim 4.3.10 implies that the smallest algebraic subvariety $W^{\boldsymbol{q}}_{\boldsymbol{g}}$ of $S(F)^n$ that is defined over L_U and contains $q(\boldsymbol{g}(\boldsymbol{x})) \in S(F)^n$ determines $\mathrm{tp}_{qf}(\boldsymbol{g}(\boldsymbol{x}))/L_U)$. Since $\mathrm{Mod}^1_{\boldsymbol{g}}$ is true in $\boldsymbol{q}$, $W^{\boldsymbol{q}}_{\boldsymbol{g}} \subseteq Z^{\boldsymbol{q}}_{\boldsymbol{g}}$. But since by Lemma 4.2.12 $Z^{\boldsymbol{q}}_{\boldsymbol{g}}$ is fixed setwise by σ (the map described after Definition 4.3.9), being defined over $E^{ab}(\Sigma)$, we have $Z^{\boldsymbol{q}'}_{\boldsymbol{g}} = Z^{\boldsymbol{q}}_{\boldsymbol{g}}$, and therefore $W^{\boldsymbol{q}'}_{\boldsymbol{g}} \subseteq Z^{\boldsymbol{q}'}_{\boldsymbol{g}}$. Now applying $\mathrm{Mod}^2_{\boldsymbol{g}}$ in $\boldsymbol{q}'$, we find the required x'. □

Having proved Claim 4.3.12, we can finish the argument. We need one more crucial piece for the 'forth'. What if $x \in D - U_D$? For this, we need $\boldsymbol{q}'$ to be ω-saturated (realize all types over finite sets).

Theorem 4.3.13. *Suppose that $\boldsymbol{q}$ and $\boldsymbol{q}'$ are ω-saturated. Then the qf-system described in Definition 4.3.9 is a back and forth*; *hence*, $T(\boldsymbol{p})$ *is complete.*

Proof. Suppose f is an isomorphism between $U \subseteq \boldsymbol{q}$ and $U' \subseteq \boldsymbol{q}'$. Then f restricts to a G-equivariant injection of U_D into $U_{D'}$ and an embedding of $S(L_U)$ into $S(F')$ induced by an embedding σ of L_u into $S(F')$, that fixes $E^{ab}(\Sigma)$.

For $x \in \boldsymbol{q} - U$, we must find $x' \in U'$ so that $f \cup (x, x')$ generates an isomorphism between the structures generated by $U \cup \{x\}$ and $U' \cup \{x'\}$. If $x \in S$, $x = q(\check{x})$ for some $\check{x} \in D$ so we restrict to that case. If $x \in U_D$, x' exists as U'_D is closed under action by G. Since the coordinates of special points are in $E^{ab}(\Sigma)$, whose points are all named, for a special point x, x' must equal x.

The difficult case is when $x \in D - U_D$ is Hodge generic. But we noted in Claim 4.3.10 that it suffices to simultaneously realize all types

$$\mathrm{tp}_{qf}((q(g_1x), \ldots, q(g_nx))/U)$$

for all $\boldsymbol{g}$ (of arbitrary length). A slight variant on the argument for Claim 4.3.12 still holds if for fixed x, we replace a single $\boldsymbol{g}$ by an arbitrary finite set of $\boldsymbol{g}$. By compactness, the entire type is consistent and so satisfied in the ω-saturated $\boldsymbol{q}'$. There is one final step. By induction we have to choose x' for a sequence $\boldsymbol{x}, \boldsymbol{y}, x$ where $\boldsymbol{x} \in U_D$ and $\boldsymbol{y} \in U_S^k$ for some k. But what if $x \in U_S$? By Claim 4.3.10, $\mathrm{tp}_{qf}(\boldsymbol{x}, \boldsymbol{y})$ is determined by $\mathrm{tp}_{qf}(\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{y})$ (in the field sort). That we can choose of $x' \in U_S'$ to satisfy $f(\mathrm{tp}_{qf}(\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{y}))$ is now immediate by ω-saturation and quantifier elimination in the field sort.

By Karp's theorem [5, Theorem 3], the existence of the back and forth implies all ω-saturated models of $T(\boldsymbol{p})$ are $L_{\omega_1,\omega}$ (indeed, $L_{\infty,\omega}$) elementarily equivalent. Every model has an ω-saturated elementary extension, so $T(\boldsymbol{p})$ is complete. $\square$

4.4. ***Galois representations and finite index conditions.*** In this section we begin by considering the action of discrete and Galois groups on the domain and field sorts. Then we unite these approaches by defining a Galois representation. We then state the key to establishing categoricity, a consequence of Serre's open mapping theorem.

4.4.1. *Two views: domain and field sort.* We explore the following diagram, which links the domain sort (via the quotient) with the field sort.

$$\begin{array}{ccc}
\mathbb{H}_{\bar h} \approx \Gamma_{\bar h} \setminus \mathbb{H} & \xrightarrow{[\phi_{\bar h}]} & Z_{\bar h} \\
{\scriptstyle \mathrm{id}_{\mathbb{H}_{\bar g}}} \big\downarrow & & \big\downarrow {\scriptstyle \psi_{\bar h, \bar g}} \\
\mathbb{H}_{\bar g} \approx \Gamma_{\bar g} \setminus \mathbb{H} & \xrightarrow{[\phi_{\bar g}]} & Z_{\bar g}
\end{array}$$

Convention 4.4.1. $\boldsymbol{g} = \langle e, g_1 \ldots g_{n-1} \rangle$ has length n. We restrict to $\boldsymbol{g}$ with $\Gamma_{\boldsymbol{g}} \trianglelefteq \Gamma$ (normal subgroup). Recall $Z_{\boldsymbol{g}} \subseteq S(\mathbb{C})^{\lg(\boldsymbol{g})}$.

We have two views of 'essentially' the same map. The first moves to a quotient on the domain side which is not τ-definable; the second 'names' the range of the first in the target side. We begin with *quotient data* but with manifestations in both the domain and target.

Domain/quotient data: The first view motivates id for identity.

Definition 4.4.2. Let $\boldsymbol{g} \subseteq \boldsymbol{h}$. Define $\mathrm{id}_{\boldsymbol{hg}} : \mathbb{H}_{\boldsymbol{h}} \to \mathbb{H}_{\boldsymbol{g}}$ by $[x]_{\Gamma_{\boldsymbol{h}}} \mapsto [x]_{\Gamma_{\boldsymbol{g}}}$.

The normality hypothesis implies that $\Gamma_{\boldsymbol{g}} / \Gamma_{\boldsymbol{h}}$ acts on $\mathbb{H}_{\boldsymbol{g}}$: for $\lambda \in \Gamma_{\boldsymbol{g}}$, $\lambda[x]_{\Gamma_{\boldsymbol{g}}} := [\lambda x]_{\Gamma_{\boldsymbol{g}}}$, so the representatives λ_i of the cosets of $\Gamma_{\boldsymbol{g}} / \Gamma_{\boldsymbol{h}}$ index the equivalence classes; thus the action is transitive.

Field data: We define the right-hand column of the diagram.

Definition 4.4.3. 1. For $\boldsymbol{g} \subseteq \boldsymbol{h}$, $\lg(\boldsymbol{g}) = n$, $\lg(\boldsymbol{h}) = m$, $\psi_{\boldsymbol{h},\boldsymbol{g}}$, denotes the restriction of the natural projection from $S(\mathbb{C})^m$ onto $S(\mathbb{C})^m$ to a map from $Z_{\boldsymbol{h}} \subseteq S(\mathbb{C})^m$ onto $Z_{\boldsymbol{g}} \subseteq S(\mathbb{C})^n$.

2. Choose $z \in Z_{\boldsymbol{g}}$ and let $L = L_z$ be a finitely generated extension of the defining field for S such that z is defined over L. Write $\bar{L}$ for $\mathrm{acl}(L)$.
3. Now, $\mathrm{Aut}(\mathbb{C}/L)$ acts on the fiber of $\psi_{\boldsymbol{h},\boldsymbol{g}}$ over z, by its action on the coordinates of z; as it would for any definable finite-to-one map from $Z^m_{\boldsymbol{h}} \to Z^n_{\boldsymbol{g}}$.

To connect the two sides, conjugating by $[\phi_{\boldsymbol{h}}]$, $\mathrm{Aut}(\bar{L}/L)$ acts on $\mathrm{id}^{-1}_{\boldsymbol{hg}}(\phi^{-1}_{\boldsymbol{g}}(z))$.

Lemma 4.4.4 [19, p. 14, top]. $\mathrm{Aut}(\mathbb{C}/L)$ *acts on the fiber of* $\psi_{\boldsymbol{h},\boldsymbol{g}}$ *over* z, (*and so via* $[\phi_{\boldsymbol{h}}]$ *on* $\mathrm{id}^{-1}_{\boldsymbol{hg}}(z)$). *This action commutes with the action of the free and transitive* (*simply transitive*) *action of* $\Gamma_{\boldsymbol{g}}/\Gamma_{\boldsymbol{h}}$ *on the fibers of* $\mathrm{id}_{\boldsymbol{h},\boldsymbol{g}}$. *Thus we have a homomorphism* (*Galois representation*) $\rho^z_{\boldsymbol{g},\boldsymbol{h}}$ *from* $\mathrm{Aut}(\bar{L}/L)$ *into* $\Gamma_{\boldsymbol{g}}/\Gamma_{\boldsymbol{h}}$.

4.4.2. *Galois representation.* While the notion of a representation of a group A frequently refers to linear representations, a homomorphism of A into a matrix group B, here we will discuss specific examples of a more general notion: *a representation of A is a homomorphism of A into a group B. This is a* ***Galois representation*** *if A is the Galois group of one field over another.* In Section 4.4.1, we gave Galois representations of $\mathrm{Aut}(\bar{L}/L)$ into $\Gamma_{\boldsymbol{g}}/\Gamma_{\boldsymbol{h}}$. In order to understand how to combine the actions of the $\Gamma_{\boldsymbol{g}}/\Gamma_{\boldsymbol{h}}$ as $\boldsymbol{g}, \boldsymbol{h}$ vary, we need the notion of inverse limit.

Definition 4.4.5 (inverse limit). Given a directed set $(I, \leq)$ an *inverse system* on I is a family of structures $\langle A_i : i \in I\rangle$, and for $i < j$, maps f_{ij} from A_j to A_i such that $i < j < k$ implies $f_{ij} \circ f_{jk} = f_{ik}$.

An *inverse limit* of this inverse system is an object $\hat{A} = \varprojlim A_i$ and a family of morphisms $g_i : \hat{A} \to A_i$ such that

(1) for all $i < j$ in I, $f_{ij} \circ g_j = g_i$, and

(2) given any A' and family g'_i satisfying (1) there is a unique morphism $h : \hat{A} \to A'$ such that for all $i \in I$, $g'_i = g_i \circ h$.

Definition 4.4.6 (Galois representations of inverse limits). We work with a modular curve $S(\mathbb{C}) = \Gamma \setminus \mathbb{H}$ which is defined over $E^{ab}(\Sigma)$ (Notation 4.3.3). Since each $\Gamma\boldsymbol{g} \subseteq \Gamma$, $\rho^z_{\boldsymbol{g},\boldsymbol{h}} : \mathrm{Aut}(\bar{L}/L) \to \Gamma$ and by taking an inverse limit of the representations $\rho^z_{\boldsymbol{g},\boldsymbol{h}}$, we obtain:

$$\rho^z : \mathrm{Gal}(\bar{L}/L) \to \bar{\Gamma}$$

where $\bar{\Gamma} = \varprojlim_{\boldsymbol{h}} \Gamma/\Gamma_{\boldsymbol{h}}$. The $\boldsymbol{h}$ range over all finite sequences as Convention 4.4.1. See Definition 4.4.5 and compare [19, p. 16].

For any groups $H_1 \leq H_2$ that act on a set X the H_1-orbits of X partition the H_2-orbits. So if $[H_2 : H_1]$ is finite and H_2 is infinite, the orbits will have the same cardinality and the smaller $[H_2 : H_1]$ is, the closer we are to an isomorphism.

Now, we can state the first of two crucial sufficient conditions for categoricity.

Definition 4.4.7 (FIC1). The *first finite index condition* (FIC1) is satisfied by a modular curve $p\colon \mathbb{H} \to S(\mathbb{C})$ if:

For any nonspecial points $x_1, \dots x_m \in \mathbb{H}$ in distinct G-orbits (Definitions 4.4.2 and 4.4.3) and for any field L containing the field over $E^{ab}(\Sigma)$ along with the coordinates of the $p(x_i)$, the image of the induced homomorphism $\rho : \mathrm{Gal}(\bar{L}/L) \to \bar{\Gamma}^m$ has finite index in $\bar{\Gamma}^m$.

Recall from Claim 4.3.10 that

$$r_d(\boldsymbol{v}) \models \mathrm{tp}_{qf}(d/U),$$

where $r_d(\boldsymbol{v}) = \bigcup_{\boldsymbol{g}\in G} \mathrm{tp}_{qf}(q(\boldsymbol{g}(d))/U) = \mathrm{tp}_{qf}(\langle q(gd) : g \in G\rangle/U)$. The argument for Claim 4.3.10 began with the observation that $r_d(\boldsymbol{v})$ implied, in particular, that $d \notin D_U$, so d is an independent Hodge generic. We will deduce from Lemma 4.4.8 that (under FIC1) only finitely many tuples $\boldsymbol{g}$ from r_d are really needed.

Lemma 4.4.8. *Assume* FIC1. *Then, for each z, for some $\hat{\boldsymbol{g}}$, the map*

$$\rho_z : \mathrm{Aut}(\bar{L}/L_{\hat{\boldsymbol{g}}}) \mapsto \bar{\Gamma}^m_{\hat{\boldsymbol{g}}} = \varprojlim_{\boldsymbol{h}\supseteq\hat{\boldsymbol{g}}} (\Gamma_{\hat{\boldsymbol{g}}}/\Gamma_{\boldsymbol{h}})^m$$

is surjective.

Proof. Let $I = \mathrm{im}(\rho_z)$ and let $k = [\bar{\Gamma} : I]$. Suppose not. Choose $\hat{\boldsymbol{g}}$ with $\boldsymbol{g} \subseteq \hat{\boldsymbol{g}}$ such that $[\Gamma_{\boldsymbol{g}} : \Gamma_{\hat{\boldsymbol{g}}}] = k$. Thus, for any $\boldsymbol{h} \supseteq \hat{\boldsymbol{g}}$, ρ_z must be onto $\Gamma_{\hat{\boldsymbol{g}}}/\Gamma_{\boldsymbol{h}}$. For, if not, there is an $\eta \in \Gamma_{\hat{\boldsymbol{g}}}/\Gamma_{\boldsymbol{h}}$ and that is not in I; it must be in a new coset of I in $\bar{\Gamma}$, contrary to the choice of $\hat{\boldsymbol{g}}$. ☐

Corollary 4.4.9. *Assume* FIC1. *For $d \in D - U$,*

$$\mathrm{tp}_{qf}(q(\hat{\boldsymbol{g}}(d))/U) \models r_d(\boldsymbol{v}) \models \mathrm{tp}_{qf}(d/U).$$

Proof. The second implication is Claim 4.3.10. For the first, choose any $\boldsymbol{h} \supseteq \hat{\boldsymbol{g}}(d)$ and let $m = \lg(\hat{\boldsymbol{g}})$, $r = \lg(\boldsymbol{h})$. Let $\mathcal{F} \subseteq Z^r_{\boldsymbol{h}}$ be the fiber over $\hat{\boldsymbol{g}}(d') \in Z^m_{\hat{\boldsymbol{g}}}$ of the finite-to-one map $\psi_{\boldsymbol{h}\hat{\boldsymbol{g}}} : Z^r_{\boldsymbol{h}} \to Z^m_{\hat{\boldsymbol{g}}}$. Similarly, $\mathrm{tp}_{qf}(\boldsymbol{h}(d)/L_U)$ is determined by the $\mathrm{Aut}(\mathbb{C}/L_U)$-orbit $\mathcal{G} \subseteq \mathcal{F}$ containing $\boldsymbol{h}(d)$. Then, $\mathrm{tp}_{qf}(\boldsymbol{h}(\boldsymbol{x})/L_U)$ is determined by the $\mathrm{Aut}(\mathbb{C}/L_U)$-orbit $\mathcal{G} \subseteq \mathcal{F}$ containing $\boldsymbol{h}(\boldsymbol{x})$. But $\mathcal{G} = \mathcal{F}$, since ρ_z induces a homomorphism from $\mathrm{Aut}(\mathbb{C}/L_U)$ *onto* $\Gamma_{\hat{\boldsymbol{g}}}/\Gamma_{\boldsymbol{h}}$ and $\Gamma_{\hat{\boldsymbol{g}}}/\Gamma_{\boldsymbol{h}}$ acts transitively on the fiber. Since this holds for any such $\boldsymbol{h}$, we finish. ☐

We turn now to the infinitary axioms that are needed to obtain categoricity.

Notation 4.4.10 (infinitary axioms).

1. Φ_∞ is the $L_{\omega_1,\omega}$ sentence asserting that for (D, S, q) both the dimension of the field bi-interpretable with S and of the strongly minimal structure $\langle D, \{f_g : g \in \Gamma\}\rangle$ are infinite.

2. *SF* (standard fibers) denotes the $L_{\omega_1,\omega}$-axiom:

$$(\forall x \forall y \in D)(q(x) = q(y) \to \bigvee_{g \in \Gamma} x = f_g(y)).$$

3. $T^\infty(\boldsymbol{p})$ denotes $\mathrm{Th}(\boldsymbol{p}) \cup \{\Phi_\infty\}$.
4. $T^\infty_{SF}(\boldsymbol{p})$ denotes $\mathrm{Th}(\boldsymbol{p}) \cup \{SF\} \cup \{\Phi_\infty\}$.

Definition 4.4.11. For $\langle D, S(F), q\rangle \models T^\infty_{SF}(\boldsymbol{p})$ and $X \subset D \cup S(F)$,

$$\mathrm{cl}(X) = q^{-1}(\mathrm{acl}(q(X)))$$

where acl is the field algebraic closure in F.

An essential consequence of the standard fibers axiom is that Definition 4.4.11 defines an almost quasiminimal closure relation satisfying the countable closure condition from Definition 3.2.4. This closure dimension restricts on the separate sorts to the dimension of the constituent strongly minimal sets that is expressed in Φ_∞. This accomplishes the aim of an ($L_{\omega_1,\omega}$-complete so $\aleph_0$-categorical) $L_{\omega_1,\omega}$ theory with arbitrarily large models.

A class $\boldsymbol{K}$ of models has $\aleph_0$-homogeneity over $\varnothing$ (Definition 3.2.5) (the precise statement is from [19, p. 4]) if the models of $\boldsymbol{K}$ are pairwise qf-back and forth equivalent (Definition 4.3.7).

Theorem 4.4.12 [15, Theorem 4.11]. *If the standard model $\boldsymbol{p}$ of a modular curve satisfies* FIC1, *then the class of models of $T^\infty_{SF}(\boldsymbol{p})$ is $\aleph_0$-homogenous over $\varnothing$. In particular, by Karp [5; 27], all models of $T^\infty_{SF}(\boldsymbol{p})$ are back and forth equivalent and so satisfy the same sentences of $L_{\omega_1,\omega}$.*

Proof. Our task is to replace the ω-saturation hypothesis from Theorem 4.3.13 by adding the infinitary axioms and the condition FIC1. As in the proof of Theorem 4.3.13 we need only worry about Hodge generic points. Suppose we have a partial function f from $\boldsymbol{q}$ to $\boldsymbol{q}'$ with domain and range U and U' as in Theorem 4.3.13 between models $\boldsymbol{q}$ and $\boldsymbol{q}'$ of $T^\infty_{SF}(\boldsymbol{p})$. Proceed as in the proof of the second paragraph of Theorem 4.3.13. We vary the argument for the 'difficult case' from the 3rd paragraph. Choose $\hat{\boldsymbol{g}}$ by Lemma 4.4.8. Taking $\hat{\boldsymbol{g}}$ for the $\boldsymbol{g}$ in Claim 4.3.12, for $x \in U_D$, there is an $x' \in U_{D'}$ such that

$$q(\hat{\boldsymbol{g}}(x')) \in S(F')^m \tag{$*$}$$

realizes $f(\mathrm{tp}_{qf}(q(\hat{\boldsymbol{g}}(x))/L_U))$. We want to show that the same choice x' satisfies $(*)$ for every $\boldsymbol{h} \supseteq \hat{\boldsymbol{g}}$. This is immediate from Corollary 4.4.9. The argument is completed by induction as in the 'final step' of the proof of Theorem 4.3.13. □

Remark 4.4.13 (FIC2). Like FIC1, FIC2 is a finite index condition on Galois representations into inverse limits. Now, however there are independence conditions over the ground field. [15, Condition 4.8] provides sufficient conditions so that a minor modification of the proof of Theorem 4.4.12, shows FIC2 implies homogeneity over models; pairs of models are back and forth equivalent over a countable submodel. *This is the first place in the argument where types over countable algebraically closed fields rather than the empty set* (*i.e., a fixed countable field*) *are encountered.* Combining this result with Theorem 4.4.12, the homogeneity conditions are now stronger than those defining quasiminimal excellence in [8]. Thus, we apply that paper and obtain:

Theorem 4.4.14. *For any modular curve interpreted as a standard model* $\boldsymbol{p}$ (Notation 4.3.3) *for* $T^{\infty}(\boldsymbol{p})$, $T^{\infty}_{SF}(\boldsymbol{p})$ *is almost quasiminimal excellent and so categorical in every infinite power.*

Proof. We need only that FIC1 and FIC2 hold for all modular curves. This is proved in [15, §5], where the proof for FIC1 relies heavily on [44, §6] and FIC2 on [43]. □

With further effort they extend this result to Shimura curves.

Remark 4.4.15. Keisler's theorem [29, Corollary 5.10] and work of Shelah [3, §7] show that an $\aleph_1$-categorical sentence ϕ of $L_{\omega_1,\omega}$ not only has only countably many types in any countable fragment of $L_{\omega_1,\omega}$ containing ϕ (Keisler) but has a completion[5] (Shelah). Equivalently, the completion must specify the isomorphism type of the countable model. The only such completion consistent with having an uncountable model is adding Φ_∞.

We have used FIC1 to prove categoricity in all powers. In fact, $\aleph_1$-categoricity implies FIC1. For this, [15; 19] argue that the weaker hypothesis of having just countably many types over the empty set in the theory T^{∞}_{SF} implies FIC1. If FIC1 holds, for some z, by Lemma 4.4.8, for every $\boldsymbol{g}$, there is $\boldsymbol{h} \supseteq \boldsymbol{g}$ with a $\Gamma_{\boldsymbol{g}}/\Gamma_{\boldsymbol{h}}$-orbit contained in $\psi^{-1}_{\boldsymbol{hg}}(z)$ that projects to that $\Gamma_{\boldsymbol{g}}$ orbit. So under the assumption that FIC1 fails, there is a $\boldsymbol{g}$, such that for every $\boldsymbol{h} \supseteq \boldsymbol{g}$ there are distinct $\Gamma_{\boldsymbol{g}}/\Gamma_{\boldsymbol{h}}$-orbits O_1, O_2 contained in $\psi^{-1}_{\boldsymbol{hg}}(z)$ that project to the same $\Gamma_{\boldsymbol{g}}$-orbit.

By Claim 4.3.10, if two points are Galois equivalent they realize the same quantifier free τ-type; so O_1, O_2 realize distinct Galois orbits (and so any two orbits that project to them must realize distinct τ-types). But since $\overline{\Gamma}$ acts transitively on each $Z_{\boldsymbol{g}}$, there is a complete tree of splittings of $\mathrm{Aut}(\mathbb{C}/L)$ orbits that all project to z. This contradicts Keisler's theorem. So $\aleph_1$-categoricity of T^{∞}_{SF} implies FIC1.

Remark 4.4.16. [15, §5], using both Serre's open mapping theorem [44, §6] for the finite index condition and work by [43] on Shimura curves show FIC1 and FIC2 hold for all modular and Shimura curves. So our remaining sections concern higher-dimensional varieties. FIC1 is known for some higher dimensional Shimura varieties and conjecturally for others, while FIC2 is true for all [19].

[5] That is, a sentence ϕ^* that implies ϕ and decides every $L_{\omega_1,\omega}$-sentence.

[15] use both to prove categoricity. Since the Galois group is not accessible in our formal language, FIC1 cannot be directly expressed in the two-sorted theory. So the goal of a 'fully formal invariant' cannot be achieved unless explicit reliance on the finite index conditions as an hypothesis is avoided.

5. First order excellence

Here is the opening paragraph of [9].

> *Let $\mathbb{G} = \mathbb{G}^n$ be a complex algebraic torus, or let $\mathbb{G}$ be a complex abelian variety. Considering $\mathbb{G}(\mathbb{C})$ as a complex Lie group, with $\boldsymbol{L}\mathbb{G} = T_0(\mathbb{G}(\mathbb{C}))$ its (abelian) Lie algebra, the exponential map provides a surjective analytic homomorphism*
>
> $$\exp : \boldsymbol{L}\mathbb{G} \twoheadrightarrow \mathbb{G}(\mathbb{C}).$$

In the spirit of Zilber, their paper aims at finding 'algebraic descriptions' of the cover exp which characterize the standard structure (at least up to categoricity in power). They solve a more general problem by providing a first order theory $\hat{T}$ for the situation and showing each model $\tilde{M}$ ($\hat{M}$ here) of $\hat{T}$ is determined by relations among two designated substructures and a certain transcendence degree. In this generality, the result is proved for any abelian group of finite Morley rank (henceforth fmr groups). Then, under slightly stronger hypotheses, the result becomes a true categoricity result for, in particular, an abelian variety defined over a number field.

We address in this section four new ingredients: formalized nonstandard covers, 'first order excellence', Kummer theory, and a distinction between classification and categoricity. First order excellence appears to be both necessary and applicable for higher order Shimura varieties.

As noted in [9], the quasiminimal approach studied earlier in this paper suffices to prove the $L_{\omega_1,\omega}$-categoricity in power for abelian varieties. The goal of this section is to identify the distinctive elements of the proof from [9] that later reappear in [19].

5.1. ***The two-sorted structure and fmr groups.*** A first order theory T is stable in κ if any $M \models T$, with $|M| = \kappa$, $|S(M)| = \kappa$. ($S(M)$ denotes the set of 1-types over M.) Morley showed that ω-stability (more properly, $\aleph_0$-stability) of a theory T is equivalent to stability in all powers (and also to the Morley rank having an ordinal value for each type). We need here a slightly weaker condition called *superstability*: T is stable in κ if $\kappa \geq 2^{\aleph_0}$.

The theory of $(\mathbb{Z}, +)$ is one of the prototypical strictly superstable theories[6] (that is, superstable, but not $\aleph_0$-stable). One can fix arbitrarily the congruence class of an element x for each n. This gives $2^{\aleph_0}$ distinct types realized by nonstandard integers.

[6]The other one is the theory of countably many equivalence relations E_n such that for each n, each E_n-class is split into infinitely many E_{n+1}-classes (and $E_{n+1} \subseteq E_n$).

There is an extensive theory of fmr groups (see [1; 12]). We need here only the basics. In particular, Macintyre's result [32] that an ω-stable group is divisible by finite. We now introduce the two-sorted theory; with that notation we are able at the end of this section to outline the main steps of the proof.

Unlike [15], where $\varprojlim Z_g$ is in the background of the proof of (our) Theorem 4.4.12 but not the statement, [9] builds the structure of nonstandard covers into the vocabulary of the two sorted structure by the ρ_n below.

Bays, Hart, and Pillay [9, §2.2] use the inverse limit of Definition 5.1.1 for *divisible* abelian groups; although it is not profinite, they refer to it as a profinite universal cover denoted $\hat{G}$ of G and G is renamed as M. Although the hat has only one meaning in [9], it becomes overloaded here so we denote the inverse limit defined below as $\tilde{M}$. While in [9] a typical 2-sorted (3-sorted in Section 8) structure $\hat{\tau}$ is represented as either $(\tilde{M}, M)$ or $\tilde{M}$, we write $\hat{M} = (\tilde{M}, M)$ and $\tilde{M}$ for the or (profinite cover) inverse limit from [9, §1.2, 2.1] as that is the actual usage in most of the cited paper.

Definition 5.1.1 ($\tilde{M}$). Given a commutative, divisible, abelian group $(M, +)$, consider the inverse limit $\tilde{M} = \varprojlim M_n$ of isomorphic copies M_m of M with the index set partially ordered by $m \le n$ if and only if $m|n$ and with maps η_{nm} (multiplication by $\frac{m}{n}$) taking $M_n \mapsto M_m$. Concretely, $(\tilde{M}, +)$ is the subgroup of the direct product of ω copies of M, containing those sequences $(\langle g_k : 1 \le k < \omega\rangle)$ such that if $k = nm$, $g_m = n \times g_k$ and $g_n = m \times g_k$.

Notation 5.1.2 (the vocabulary $\hat{\tau}$). Let $\mathbb{G}$ be the given abelian group and $T := \mathrm{Th}(\mathbb{G})$ in a large enough countable language that T has quantifier elimination. Further, let $\hat{T}$ be the theory of $(\hat{\mathbb{G}}, \mathbb{G})$ in the two-sorted language $\hat{\tau}$ consisting of the maps $\rho_n : \hat{\mathbb{G}} \to \mathbb{G}$ for each n, the theory T and, for each $\mathrm{acl}^{eq}(\varnothing)$-definable subgroup H of G, a predicate H for H and a predicate $\hat{H}$ for $\{x \in \hat{G} : \rho_n(x) \in H, n \in \mathbb{N}\}$.

Although the kernel of $\rho = \rho_1$ is definable in the vocabulary given, a further predicate $\ker^0$ is included denoting the divisible part of the kernel (otherwise, it is only type-definable).

The axioms [9, §2.5] of $\hat{T}$ are chosen so as to ensure the next result holds.

Theorem 5.1.3 [9, §2.7, 2.8, 2.21]. *For an fmr group $\mathbb{G}$, $(\varprojlim \mathbb{G}, \mathbb{G}, \rho_0) \models \hat{T}$ and therefore $\hat{T}$ admits quantifier elimination and is superstable of finite U-rank.*

Although the T in Notation 5.1.2 is ω-stable, $\hat{T}$ is only superstable; also, many elements of $\ker(\rho)$ are not divisible in $\ker(\rho)$.

Remark 5.1.4 (quasiminimality, unidimensionality, notop). Abelian varieties as opposed to fmr groups, can be handled either by the quasiminimality methods of Section 4 or by the methods described in this section. A crucial distinction from

Section 4 is that the former considered only the theory of unary functions from a group acting on the domain, while here we have the full group structure.

To explain the fmr proof we need some further model theoretic background. In general two types p, q over M are *orthogonal* when in different models N extending M the number of realizations of p and q can be varied arbitrarily. *Non-orthogonality* for strongly minimal sets has a particularly clear meaning. The strongly minimal sets D_1 and D_2 are nonorthogonal if there is a definable finite to finite binary relation on $D_1 \times D_2$. A theory is unidimensional if all types are nonorthogonal.

The three features that underlie the [9] proof are:

1. A fmr abelian group has *finite width* [2, XV.1] (aka almost $\aleph_1$-categorical [31]): Any model is the algebraic closure of the union of the bases of a collection of strongly minimal D_i for $i < n < \omega$. The D_i are defined over the prime model (the unique up to isomorphism model elementarily embedded in every model of the theory).
2. In models of $\hat{T}$ with M_0 the prime model of T and where $\mathbb{G}$ is defined over a number field k_0, Kummer theory allows the control of $\rho^{-1}(M_0)$ by the kernel $\rho^{-1}(0)$.
3. In studying abelian varieties the n in 1) can be taken as 1 because the variety is interalgebraic with an algebraically closed field and so *almost strongly minimal* ($M = \mathrm{acl}(D)$ for strongly minimal D).

Since Kummer theory doesn't apply to arbitrary Shimura varieties, features 2 and 3 fail for more general higher-dimensional Shimura varieties (see Section 6).

5.2. ***First order excellence and fmr groups.*** Shelah's main gap program defines a sequence of properties X of countable first order theories forming a sequence of dichotomies [4, §5.5] such that: if T satisfies X, T has the maximal number of models in every uncountable cardinal. If T fails X, the models of T satisfy conditions useful for classification. (e.g., stability implies the existence of the 'nonforking' independence relation). The positive side of the final dichotomy in the sequence is superstable *without the omitting types order property* (denoted *notop*). Under this hypothesis, Shelah ([47] and earlier papers) showed that an appropriate class of models of T had a notion of independence among structures with n-amalgamation for all n that yields the classification of models. Hart [26] reduced the amalgamation requirement to 2-amalgamation and this reduction was extended to the quasiminimal excellent case in [8]. In Section 6, we note this 'notop' approach is used to study higher-dimensional Shimura varieties.

In Section 3 of [9] the techniques of [26] are adapted to the specific framework here to establish a decomposition of models of $\hat{T}$ analogous to that in Remark 5.1.4 for models of T. This yields

Theorem 5.2.1 [9, Theorem 3.31]. *Each model $\hat{\mathbb{M}}$ of $\hat{T}$ is determined up to isomorphism by the transcendence degree of the algebraically closed field K such that $M \cong \mathbb{G}(K)$, the isomorphism type of the inverse image, $\hat{M}_0$, of the prime model M_0 of T, and the isomorphism type of M over M_0.*

5.3. ***Abelian varieties.*** From the model theoretic standpoint, an *abelian variety* is a complete algebraic variety whose points form a group such that the group operations are definable in the ambient field. For abelian varieties, Kummer theory eliminates (as in [7; 22]) the reliance in Theorem 5.2.1 on knowing the isomorphism type of $\check{M}_0$ over the kernel. The situation described in the opening paragraph of Section 5 is a special case. Namely, let $\mathbb{G}$ be (the formula defining) an abelian variety $\mathbb{G}(K)$ over a field K as in the introduction to Section 5. Assume $\mathbb{G}(\mathbb{C})$ and its ring of endomorphisms are definable over a number field k_0. With this notation:

Theorem 5.3.1 [9, Theorem 4.6]. *a model $\hat{M} = \langle \tilde{M}, M, q \rangle$ of $\hat{T}$ is determined up to isomorphism by the transcendence degree of the algebraically closed field K such that $M \cong \mathbb{G}(K)$, and the $\hat{\tau}$ isomorphism type of* $\ker \rho$.

Remark 5.3.2 (complete formal invariant). Theorem 5.3.1 gives categoricity in all uncountable cardinalities by adding the $L_{\omega_1,\omega}$ sentence characterizing the standard kernel. But Theorem 5.3.1 is more general than categoricity; it shows that models with nonstandard (possibly uncountable) kernel are characterized by the $\hat{\tau}$-diagram of the kernel. Of course, this statement cannot be formalized in languages with bounded length of conjunctions since the kernels can be arbitrarily large. But Zilber's goal (just after Notation 1.0.1) only aimed at complete formal characterization for prototypical mathematical structures.

6. Higher-dimensional Shimura varieties

A Shimura variety is a higher-dimensional generalization of a modular curve that arises as a quotient variety of a Hermitian symmetric space X^+ by a congruence subgroup of a reductive algebraic group defined over $\mathbb{Q}$. We consider Shimura varieties that are moduli spaces for generalized algebraic varieties. Rather than discussing further technical details on the definition of a Shimura datum (G, X), we survey the differences that arise in generalizing the results in Remark 4.4.16 about Shimura curves to higher-dimensional Shimura varieties: $S(\mathbb{C}) = \Gamma \setminus X^+$.

Central difficulties arise directly from the higher dimension in two ways. First, in the curve case the 2-sorted structure is (almost)-quasiminimal because the variety in field sort is a curve and so strongly minimal and the geometric closure on the cover sort is given by $a \in \mathrm{cl}(X)$ if $a \in q^{-1}(\mathrm{acl}((q(X))$. Quasiminimality can fail in higher dimensions. Second, rather than special points which are fixed points of some g, one must treat *special subvarieties* [19, §3.4] and finite unions thereof, *special domains*. The fact that these are not merely points leads to several difficulties.

1. The structure of the covering sort is no longer strongly minimal. Even after naming the elements of the group the special subvarieties give a complicated structure on the covering sort.
2. In the curve case the intersection of special domains was a point; that may fail in higher dimensions.
3. The theories of two inverse limit structures $\hat{\boldsymbol{p}}$ and $\tilde{\boldsymbol{p}}$ are considered as the covering space. The first structure is the analog of $\varprojlim Z_{\boldsymbol{g}}$ (Notation 4.2.11). The second consists only of the standard points of this limit. The canonical universal cover $\boldsymbol{p}$ satisfies the first order $\mathrm{Th}(\tilde{\boldsymbol{p}})$ but not in general $\mathrm{Th}(\hat{\boldsymbol{p}})$ [20, Example 5.7, Corollary 5.14].
4. An $L_{\omega_1,\omega}$ categorical axiomatization is not claimed. Each model can be precisely characterized but the characterization is not in $L_{\omega_1,\omega}$. See Remark 5.3.2.
5. Finally, even this characterization depends on whether the variety under consideration satisfies *finite index conditions* as in the modular case. Although FIC1 and FIC2 are true in the modular curve case, here the truth of FIC1 for $\boldsymbol{p}$ is actually equivalent to the characterizability of models of $T_{SF}^{\inf}(\boldsymbol{p})$ since [19] shows FIC2 is true.

7. Model theory and analysis

One can signal three different model theoretic approaches to analysis:

1. *Axiomatic analysis* studies behavior of fields of functions with operators but *without* explicit attention in the formalism of continuity but rather to the algebraic properties of the functions. The function symbols of the vocabulary act on the functions being studied; the functions are elements of the domain of the model.

   Example: DCF_0 as discussed below.
2. *Definable analysis* has a lower level of abstraction; the domain of the functions remains the universe of the model. The functions being studied are the compositions of the functions named in the vocabulary; one cannot quantify over them.

   Example: *o*-minimality.
3. *Implicit analysis.* Attempts to provide 'algebraic characterizations of important mathematical structure by axiomatizations in infinitary logic that are categorical in power. Example: the material in this paper.

The first two are discussed in [4, §6.3]. The work expounded in this paper has many commonalities with a prime example of axiomatic analysis: the study of transcendence results for solutions of differential equations by the study of the ω-stable theory DCF_0 of differentially closed fields of characteristic zero. The notion of 'not integrable by elementary functions (Painlevé said 'irreducible') is formalized by 'the solution set is strongly minimal' [37]. The study of Schwartzian equations provides a general framework in which the j-function and modular curves are explored. The work includes, variations on the Ax–Lindemann–Weierstrass theorem, proofs that Generic differential equations are strongly minimal [18] and Differential Chow Varieties are Kolchin-constructible [21], and analysis of strongly minimal solution sets defined by differential equations in terms of the Zilber trichotomy and $\aleph_0$-categoricity.

But while the mathematical topics are the same, the aims are different: The covers project tries to assign a categorical description of each cover. The DCF_0 approach tries to understand transcendence results for solutions of the differential equations.

The crucial methodological difference is the two-sorted nature of the cover program. The axiomatic analysis framework is preserved in that there is no explicit treatment of convergence or continuity. But connecting the domain and target by quotients under an explicit group action as well as the use of infinitary logic provides tools not available in the earlier examples of axiomatic analysis.

8. Families of covers of algebraic curves

In recent work, Daw and Zilber [16; 17] deal with *families* of covers of curves. They build on earlier constructions we have discussed in this paper. Rather than a cover of a *single* variety, albeit one that parametrized a family of varieties, an entire family of such covers is studied and the covering space becomes an *analytic Zariski structure* [56]. In [57], the analysis of families is generalized by being placed in a geometric algebraic setting.

The most salient difference between these works and those discussed earlier in this paper is that, rather than a cover of a *single* variety, an *entire family of covers* is now the main subject. Our earlier Definition 4.2.9 is now replaced by a basic vocabulary consisting of *three* sorts, together with maps $\Gamma_N \setminus \mathbb{H} \mapsto \mathbb{C}$ covering a family of curves $S_N(\mathbb{C})$.

8.1. ***Pseudo-analytic covers of modular curves.*** Major differences of the paper [16] from the earlier discussion of modular curves include:

1. The basic vocabulary is now 3-sorted. More specifically, [16] considers structures $(D, G, j_N, \mathbb{C})$ where the $j_N : \mathbb{H} \twoheadrightarrow S_N(\mathbb{C})$. The discrete group is now given as a third sort incorporating *a group operation* (so its pregeometry is

locally modular, rather than trivial). This sort contains group with distinguished subsets[7] $(\mathrm{GL}_2^+(\mathbb{Q}), \times, \mathrm{Sl}_2(\mathbb{Z}), E(\mathbb{Q}), \{\boldsymbol{d}_q, \boldsymbol{d}'_q : q \in \mathbb{Q})\}$, where E is the collection of elliptic elements of the group; those that have unique fixed points. This structure is specified up to isomorphism by a sentence of $L_{\omega_1,\omega}$. But not all group elements are still named in the formal language.

2. The uniformizing functions j_N each map into $\boldsymbol{P}^3(\mathbb{C})$ rather than into the arbitrarily high-dimensional spaces of the maps $[\phi_{\boldsymbol{g}}]$ in [15; 19]. Furthermore, these are now defined over $\mathbb{Q}$ rather than over $E^{ab}(\Sigma)$.
3. As well as an almost quasiminimal axiomatization of the 3-sorted structure, the domain is considered as a Zariski Analytic set with a quasiminimal geometry. Both of these structures are shown to be uncountably categorical.
4. The special points *are not named*. However as in Definition 4.3.5 they are uniquely associated with elliptic elements of the group.

In many ways, this last distinction is the most important for the general program, as naming of the special points trivializes some of the arithmetic. In [16], the structure of the *family* is proved to be categorical in all uncountable cardinalities.

8.2. *Locally o-minimal covers of algebraic varieties.* The paper [57] takes a *more general* approach. It abstracts away from naming all elements of the discrete groups as earlier in this paper. The relations among the universal and finite covers are given more abstractly as properties of maps from a domain (whose smoothness is defined topologically and geometrically but not algebraically) onto families of algebraic varieties. This smoothness as well as the eventual quasiminimality for curves[8] is controlled by *external* o-minimal structures.

Remark 8.2.1. 1. The formalization is new. For a fixed model R of the theory T of a fixed o-minimal expansion of the reals (e.g the restricted analytic functions) a structure $\mathbb{U}(\mathsf{R})$ is defined. The resulting structure $\mathbb{U}(\mathsf{R})$ is an abstract Zariski structure[9].

2. Generalizing the last paragraph of Section 8.1, in the standard model the domain is a complex manifold $\mathbb{U}(\mathbb{C})$ with holomorphic maps f_i onto algebraic varieties $X_i(\mathbb{C})$ with natural projections $\mathrm{pr}_{i,j}$ among the X_i. These analytic properties are definable using theory of K-analytic sets in o-minimal expansions of the reals developed in [39; 40]. We fix $k \subseteq \mathbb{C}$ a subfield over which the varieties $\mathbb{X}_i$ are all defined.

[7] E is the elliptic Möbius transformations and the $\boldsymbol{d_q}, \boldsymbol{d_q}$ are specific *diagonal* matrices.

[8] The set-up is for arbitrary algebraic varieties, but the categoricity result is only for curves and we restrict to that case.

[9] Actually, $\mathbb{U}(\mathsf{R}) = U(K)$ where K is taken as an algebraically closed field $\mathsf{R} + i\mathsf{R}$ and $U(R)$ is constructed analogously to $U(\mathbb{C})$.

3. The ostensibly two-sorted structure of 1) becomes one-sorted because the field can be interpreted in the abstract Zariski structure. And the third sort of Section 8.1 has disappeared because the group is no longer referenced directly.
4. The o-minimal geometry of algebraic closure in $\mathbb{U}(\mathsf{R})$ imposes the desired quasiminimal geometry on $\mathbb{U}(\mathsf{R})$. The dimension function is denoted cdim for 'combinatorial dimension'. Note that the ordering is not externally imposed on $\mathbb{U}$: rather, it is implied by the predicates described in (1) above and the dimension just mentioned.
5. As before, there is an $L_{\omega_1,\omega}$ sentence that axiomatizes the quasiminimal (excellent) geometry and whose models form an AEC that is categorical in all cardinalities.

Zilber [57] proved the following theorem:

Theorem 8.2.2 (categoricity of families of smooth complex algebraic varieties [57]). *Let $\mathbb{U}$ be a cover of a family of smooth complex algebraic variety, formalized as in Remark 8.2.1, and let $\mathfrak{U}(\mathsf{R})$ be its associated $L_{\omega_1\omega}$-definable class. If* $\dim_{\mathbb{C}}(\mathbb{U}) = 1$, (*i.e.*, *if the varieties are curves*) *and* $\mathrm{cdim}(\mathsf{R}/k)$ *is infinite*, *then* $\mathfrak{U}(\mathsf{R})$ *is categorical in all uncountable cardinals.*

Zilber remarks that in the case of higher-dimensional varieties, categoricity in $\aleph_1$ can still be proved.

Example 8.2.3. Here are some examples from [57]. Fix the o-minimal expansion $\mathbb{R}_{\mathrm{An}} = \mathbb{R}_{\exp,an}$ of the reals with the exponential function and the restricted (to bounded intervals) analytic functions.

- Let $I = \mathbb{N}$, $\mathbb{U} = \mathbb{C}$, $f_k(z) = \exp(\frac{z}{k})$, $D_n = \{z \in \mathbb{C} : -2\pi n < \mathrm{Im}(z) < 2\pi n\}$. These are easily seen to provide a cover system.
- The j-function with variants j_N as uniformizers for the modular curves $\Gamma_N \setminus \mathbb{H}$ are examples; this study allows one to formalize their analytic properties in terms of o-minimality. Finally, other examples include the Siegel half-space and polarized algebraic varieties (these last examples are claimed but not developed by Zilber).

Acknowledgements

The authors want to thank many people who helped this project go through. Among them, hoping not to forget important people, are, most notably Sebastián Eterović, Jim Freitag, Jonathan Kirby, Anatoly Libgober, Ronnie Nagloo and Boris Zilber. Without their attention to our discussions (online, at conferences and on campus), this project would have been much harder to complete. Baldwin especially wants to thank Ronnie and Sebastián for hours of conversation. Villaveces especially thanks

Alex Cruz and Leonardo Cano for many helpful discussions related to these subjects in the Bogotá seminar before this project started. Finally, discussions with Thomas Kucera and Martin Bays were very important at earlier stages of the construction of this paper. Finally, the referee reports were invaluable.

References

[1] T. Altinel, A. V. Borovik, and G. Cherlin, *Simple groups of finite Morley rank*, Mathematical Surveys and Monographs **145**, American Mathematical Society, Providence, RI, 2008. MR Zbl

[2] J. T. Baldwin, *Fundamentals of stability theory*, Springer, Berlin, 1988. MR Zbl

[3] J. T. Baldwin, *Categoricity*, University Lecture Series **50**, American Mathematical Society, Providence, RI, 2009. MR Zbl

[4] J. T. Baldwin, *Model theory and the philosophy of mathematical practice: Formalization without foundationalism*, Cambridge University Press, 2018. MR Zbl

[5] J. Barwise, "Back and forth through infinitary logic", pp. 5–34 in *Studies in model theory*, edited by M. D. Morley, MAA Stud. Math. **8**, Mathematical Association of America, Buffalo, NY, 1973. MR Zbl

[6] M. Bays and B. Zilber, "Covers of multiplicative groups of algebraically closed fields of arbitrary characteristic", *Bull. Lond. Math. Soc.* **43**:4 (2011), 689–702. MR Zbl

[7] M. Bays, M. Gavrilovich, and M. Hils, "Some definability results in abstract Kummer theory", *Int. Math. Res. Not.* **2014**:14 (2014), 3975–4000. MR Zbl

[8] M. Bays, B. Hart, T. Hyttinen, M. Kesälä, and J. Kirby, "Quasiminimal structures and excellence", *Bull. Lond. Math. Soc.* **46**:1 (2014), 155–163. MR Zbl

[9] M. Bays, B. Hart, and A. Pillay, "Universal covers of commutative finite Morley rank groups", *J. Inst. Math. Jussieu* **19**:3 (2020), 767–799. MR Zbl

[10] D. Ben-Zvi, "Moduli spaces", pp. 408–419 in *The Princeton companion to mathematics*, edited by T. Gowers et al., Princeton University Press, 2008. Zbl

[11] A. Berenstein, T. Hyttinen, and A. Villaveces, "Hilbert spaces with generic predicates", *Rev. Colombiana Mat.* **52**:1 (2018), 107–130. MR Zbl

[12] A. Borovik and A. Nesin, *Groups of finite Morley rank*, Oxford Logic Guides **26**, Oxford University Press, New York, 1994. MR Zbl

[13] E. Breuillard, A. Martin-Pizarro, K. Tent, and F. Wagner, "Model theory: Groups, geometries and combinatorics", *Oberwolfach Rep.* **17**:1 (2020), 91–142. MR Zbl

[14] J. A. Cruz Morales, A. Villaveces, and B. Zilber, "Around logical perfection", *Theoria* **87**:4 (2021), 971–985. MR Zbl

[15] C. Daw and A. Harris, "Categoricity of modular and Shimura curves", *J. Inst. Math. Jussieu* **16**:5 (2017), 1075–1101. MR Zbl

[16] C. Daw and B. I. Zilber, "Modular curves and their pseudo-analytic cover", preprint, 2021. Zbl arXiv 2107.11110

[17] C. Daw and B. I. Zilber, "Canonical models of modular curves and galois action on CM-points", preprint, 2022. Zbl arXiv 2106.06387

[18] M. DeVilbiss and J. Freitag, "Generic differential equations are strongly minimal", *Compos. Math.* **159**:7 (2023), 1387–1412. MR Zbl

[19] S. Eterović, "Categoricity of Shimura varieties", preprint, 2018. arXiv 1803.06700v2

[20] S. Eterović, "Categoricity of Shimura varieties", 2022. A revision of [19] that has not circulated widely.

[21] J. Freitag, W. Li, and T. Scanlon, "Differential chow varieties exist", *J. Lond. Math. Soc.* (2) **95**:1 (2017), 128–156. MR Zbl

[22] M. Gavrilovich, "A remark on transitivity of Galois action on the set of uniquely divisible abelian extensions in $\mathrm{Ext}^1(E(\overline{\mathbb{Q}}), \Lambda)$", *K-Theory* **38**:2 (2008), 135–152. MR Zbl

[23] R. Grossberg, "Classification theory for abstract elementary classes", pp. 165–204 in *Logic and algebra*, edited by Y. Zhang, Contemp. Math. **302**, American Mathematical Society, Providence, RI, 2002. MR Zbl

[24] J. Hadamard, *The psychology of invention in the mathematical field*, Princeton University Press, 1945. Reprinted Dover, 1954. MR Zbl

[25] A. Harris, *Categoricity and covering spaces*, Ph.D. thesis, University of Oxford, 2013, available at https://www.proquest.com/docview/1683347772. MR Zbl

[26] B. Hart, "An exposition of OTOP", pp. 107–126 in *Classification theory* (Chicago, IL, 1985), edited by J. T. Baldwin, Lecture Notes in Math. **1292**, Springer, Berlin, 1987. MR Zbl

[27] C. R. Karp, *Languages with expressions of infinite length*, North-Holland, Amsterdam, 1964. MR Zbl

[28] S. Katok, *Fuchsian groups*, University of Chicago Press, 1992. MR Zbl

[29] H. J. Keisler, "Logic with the quantifier 'there exist uncountably many'", *Ann. Math. Logic* **1** (1970), 1–93. MR Zbl

[30] J. Kirby, "On quasiminimal excellent classes", *J. Symbolic Logic* **75**:2 (2010), 551–564. MR Zbl

[31] D. Lascar, "Les groupes ω-stables de rang fini", *Trans. Amer. Math. Soc.* **292**:2 (1985), 451–462. MR Zbl

[32] A. Macintyre, "On ω_1-categorical theories of abelian groups", *Fund. Math.* **70**:3 (1971), 253–270. MR Zbl

[33] D. Marker, *Model theory: An introduction*, Graduate Texts in Mathematics **217**, Springer, New York, 2002. MR Zbl

[34] J. S. Milne, "What is . . . a Shimura variety?", *Notices Amer. Math. Soc.* **59**:11 (2012), 1560–1561. MR Zbl

[35] T. Miyake, *Modular forms*, Springer, Berlin, 1989. MR Zbl

[36] M. Morley, "Categoricity in power", *Trans. Amer. Math. Soc.* **114** (1965), 514–538. MR Zbl

[37] J. C. R. Nagloo, *Model theory, algebra and differential equations*, Ph.D. thesis, University of Leeds, 2014, available at https://www.proquest.com/docview/1687708188. MR Zbl

[38] Y. Peterzil and S. Starchenko, "A trichotomy theorem for o-minimal structures", *Proc. London Math. Soc.* (3) **77**:3 (1998), 481–523. MR Zbl

[39] Y. Peterzil and S. Starchenko, "Complex analytic geometry in a nonstandard setting", pp. 117–165 in *Model theory with applications to algebra and analysis*, *vol.* 1, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **349**, Cambridge University Press, 2008. MR Zbl

[40] Y. Peterzil and S. Starchenko, "Tame complex analysis and o-minimality", pp. 58–81 in *Proceedings of the International Congress of Mathematicians*, vol. 2, edited by R. Bhatia et al., Hindustan Book Agency, New Delhi, 2010. MR Zbl

[41] B. Poizat, *Cours de théorie des modèles*, Bruno Poizat, Lyon, 1985. MR Zbl

[42] B. Poizat, *Groupes stables*, Nur al-Mantiq wal-Ma'rifah **2**, Bruno Poizat, Lyon, 1987. MR Zbl

[43] K. A. Ribet, "On l-adic representations attached to modular forms", *Invent. Math.* **28** (1975), 245–275. MR Zbl

[44] J.-P. Serre, "Propriétés galoisiennes des points d'ordre fini des courbes elliptiques", *Invent. Math.* **15**:4 (1972), 259–331. MR Zbl

[45] S. Shelah, "Classification theory for non-elementary classes, I: The number of uncountable models of $\psi \in L_{\omega_1,\omega}$, Part A", *Israel J. Math.* **46**:3 (1983), 212–240. MR Zbl

[46] S. Shelah, "Classification theory for non-elementary classes, I: The number of uncountable models of $\psi \in L_{\omega_1,\omega}$, Part B", *Israel J. Math.* **46**:4 (1983), 241–273. MR Zbl

[47] S. Shelah, *Classification theory and the number of nonisomorphic models*, 2nd ed., Studies in Logic and the Foundations of Mathematics **92**, North-Holland, Amsterdam, 1990. MR Zbl

[48] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Kanô Memorial Lectures **1**, Iwanami Shoten Publishers, Tokyo, 1971. MR Zbl

[49] K. Tent and M. Ziegler, *A course in model theory*, Lecture Notes in Logic **40**, Cambridge University Press, 2012. MR Zbl

[50] S. Vasey, "Quasiminimal abstract elementary classes", *Arch. Math. Logic* **57**:3-4 (2018), 299–315. MR Zbl

[51] B. I. Zilber, "The structure of models of uncountably categorical theories", pp. 359–368 in *Proceedings of the International Congress of Mathematicians, vol.* 1–2 (Warsaw, 1983), edited by Z. Ciesielski and C. Olech, Polish Scientific Publishers, Warsaw, 1984. MR Zbl

[52] B. Zilber, "Analytic and pseudo-analytic structures", pp. 392–408 in *Logic colloquium* (Paris, 2000), edited by S. T. R. Cori, Alexander Razborov and C. Wood, Lect. Notes Log. **19**, Assoc. Symbol. Logic, Urbana, IL, 2005. MR Zbl

[53] B. Zilber, "A categoricity theorem for quasi-minimal excellent classes", pp. 297–306 in *Logic and its applications*, edited by A. Blass and Y. Zhang, Contemp. Math. **380**, American Mathematical Society, Providence, RI, 2005. MR Zbl

[54] B. Zilber, "Pseudo-exponentiation on algebraically closed fields of characteristic zero", *Ann. Pure Appl. Logic* **132**:1 (2005), 67–95. MR Zbl

[55] B. Zilber, "Covers of the multiplicative group of an algebraically closed field of characteristic zero", *J. London Math. Soc.* (2) **74**:1 (2006), 41–58. MR Zbl

[56] B. Zilber, *Zariski geometries: Geometry from the logician's point of view*, London Mathematical Society Lecture Note Series **360**, Cambridge University Press, 2010. MR Zbl

[57] B. Zilber, "Nonelementary categoricity and projective locally o-minimal classes", *Model Theory* **3**:1 (2024), 101–117. MR Zbl

Received 8 Feb 2023. Revised 7 Sep 2023.

JOHN T. BALDWIN:
jbaldwin@uic.edu
Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL, United States

ANDRÉS VILLAVECES:
avillavecesn@unal.edu.co
Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, Colombia

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.685

Positive characteristic Ax–Schanuel

Piotr Kowalski

This expository paper is written in celebration of Boris Zilber's 75th birthday. We discuss Ax–Schanuel type statements focusing on the case of positive characteristic.

1. Introduction

During the Spring 2005 Isaac Newton Institute program "Model Theory and Applications to Algebra and Analysis" in Cambridge, I learnt that I would be a MODNET postdoc with Boris Zilber in Oxford for the academic year 2005/06. Still in Cambridge, Boris suggested that I start thinking on "positive characteristic versions of Ax's theorem". In this expository paper, I will describe what has happened next.

It may be a good moment for a general disclaimer. This is an expository paper representing my experience with respect to Boris's suggestion above and I do not claim that this paper describes adequately the state of the art in the vast area of Ax–Schanuel type problems. In particular, comparatively very little will be said about the amazing developments of Jonathan Pila (and many others) regarding the modular version of Ax–Schanuel and its applications to diophantine problems, most notably the André–Oort conjecture. I will write more about it in Section 2.

This paper is organized as follows. In Section 2, we describe the history of this circle of problems in the case of characteristic 0. In Section 3, we focus on the positive characteristic case and present some of the results I obtained following Boris's suggestion. In Section 4, we speculate on some recent ideas regarding general forms of Ax–Schanuel and its Hasse–Schmidt differential versions.

2. Characteristic zero

In this section, we summarize the characteristic 0 situation regarding the Ax–Schanuel problems. The disclaimer from the introduction applies mostly here.

Supported by the Narodowe Centrum Nauki grant no. 2021/43/B/ST1/00405 and by the Tübitak 1001 grant no. 119F397.
I would like to thank Jakub Gogolok for his comments and to the referee for the very careful and useful report.
MSC2020: primary 11J91; secondary 12H05.
Keywords: Ax–Schanuel conjecture, formal maps, transcendence.

2A. *Results.* In the 1960s, Schanuel formulated two conjectures [Lang 1966, pages 30–31]: one about transcendence of complex numbers [Ax 1971, (S)] and one about transcendence of power series [Ax 1971, (SP)]. We state them below.

Schanuel's conjecture (complex numbers). *Let $x_1, \dots, x_n \in \mathbb{C}$ be linearly independent over $\mathbb{Q}$. Then*

$$\operatorname{trdeg}_{\mathbb{Q}}(x_1, \dots, x_n, e^{x_1}, \dots, e^{x_n}) \geqslant n.$$

Schanuel's conjecture (power series). *Let $x_1, \dots, x_n \in X\mathbb{C}[[X]]$ be linearly independent over $\mathbb{Q}$. Then*

$$\operatorname{trdeg}_{\mathbb{C}(X)}(x_1, \dots, x_n, e^{x_1}, \dots, e^{x_n}) \geqslant n.$$

The conjecture on the complex numbers is open even for $n = 2$, since (using Euler's identity $e^{i\pi} + 1 = 0$) it covers the open problem of algebraic independence of π and e and it is even still unknown whether $\pi + e$ is irrational (it is named a "candidate for the most embarrassing transcendence question in characteristic zero" in [Brownawell 1998])! Schanuel's conjecture for power series was proved in [Ax 1971, (SP)].

Ax [1971] also showed the following differential version of the power series conjecture, which he actually used to show the other statements from [Ax 1971].

Differential Ax–Schanuel theorem [Ax 1971, (SD)]. *Let (K, ∂) be a differential field of characteristic 0 and C be its field of constants. For $x_1, \dots, x_n \in K$ and $y_1, \dots, y_n \in K^*$, if*

$$\partial x_1 = \frac{\partial y_1}{y_1}, \dots, \partial x_n = \frac{\partial y_n}{y_n}$$

and $\partial x_1, \dots, \partial x_n$ are $\mathbb{Q}$-linearly independent, then

$$\operatorname{trdeg}_C(x_1, \dots, x_n, y_1, \dots, y_n) \geqslant n + 1.$$

Remark 2.1. There are the following passages between the power series and the differential version of Ax's theorem above.

(1) Since the ring of power series has a natural differential structure, the differential version implies the power series version.

(2) Going the other way is more subtle. Seidenberg's embedding theorem [1958] says that any finitely generated differential field of characteristic 0 differentially embeds into the differential field of meromorphic functions on an open subset of $\mathbb{C}$. Using this theorem, one can reduce the differential version of Ax's theorem to the power series one (this is explained in detail around [Freitag and Scanlon 2018, Theorem 4.1] and in [Pila and Tsimerman 2016, Section 2.5]).

Similar passages apply to the more complicated cases of analytic (or formal) Ax–Schanuel statements versus the differential ones as well. Such more complicated cases are described below.

In a subsequent paper written one year later, Ax [1972] proved the following general geometric result about the dimension of intersections of algebraic subvarieties of complex algebraic groups with analytic subgroups.

Ax's theorem on the dimension of intersections [Ax 1972, Theorem 1]. *Let G be an algebraic group over the field of complex numbers $\mathbb{C}$. Let $\mathcal{A}$ be a complex analytic subgroup of $G(\mathbb{C})$ and V be an irreducible algebraic subvariety of G over $\mathbb{C}$. We assume that $\mathcal{K} := \mathcal{A} \cap V(\mathbb{C})$ is Zariski dense in $V(\mathbb{C})$. Then there is an analytic subgroup $\mathcal{B} \subseteq G(\mathbb{C})$ containing $V(\mathbb{C})$ and $\mathcal{A}$ such that*

$$\dim(\mathcal{B}) \leqslant \dim(\mathcal{A}) + \dim(V) - \dim(\mathcal{K}).$$

This theorem implies Schanuel's conjecture on power series by taking:

- G as the product of the vector group $\mathbb{G}_{\mathrm{a}}^n$ and the torus $\mathbb{G}_{\mathrm{m}}^n$,
- $\mathcal{A}$ as the n-th Cartesian power of the graph of the exponential map,
- V as the algebraic locus of the tuple $(x_1, \ldots, x_n, e^{x_1}, \ldots, e^{x_n})$.

Ax's theorem on the dimension of intersections applies also (more generally) to the case of the exponential map on a semiabelian variety [Ax 1972, Theorem 3]. The consequences of Ax's theorem on the dimension of intersections go beyond the case of the exponential map; for example, this theorem applies to the case of analytic maps between the multiplicative group and an elliptic curve. We state it precisely below, since this statement is amenable for a possible transfer to the positive characteristic case (see Remark 2.4).

Theorem 2.2. *Let*

$$\gamma : \mathbb{G}_{\mathrm{m}}(\mathbb{C}) \to E(\mathbb{C})$$

be an analytic epimorphism, where E is an elliptic curve. Let

$$x_1, \ldots, x_n \in 1 + X\mathbb{C}[[X]]$$

be multiplicatively independent. Then

$$\operatorname{trdeg}_{\mathbb{C}(X)}(x_1, \ldots, x_n, \gamma(x_1), \ldots, \gamma(x_n)) \geqslant n.$$

After Ax's work in 1970s, Brownawell and Kubota [1977] proved a version of the differential Ax's theorem in the case of elliptic curves, and then Kirby [2009] generalized it to arbitrary semiabelian varieties. These results were not included in [Ax 1972], however they are closely related using the "passages" from Remark 2.1. Bertrand [2008] extended [Ax 1972, Theorem 3] to commutative algebraic groups not having vector quotients (e.g., maximal nonsplit vectorial extensions of a semiabelian variety).

The differential Ax's theorem [Ax 1971, (SD)] is generalized further to "very nonalgebraic formal maps" in [Kowalski 2008, Theorem 5.5]. This generalization includes a differential version of Bertrand's result and a differential Ax–Schanuel type result about raising to nonalgebraic powers on an algebraic torus [Kowalski 2008, Theorem 6.12]. We state it below in the power series case (see Remark 2.1), since this statement has a positive characteristic interpretation (see Remark 2.4). Before the statement, we note that for $x \in 1 + X\mathbb{C}[\![X]\!]$ and $\alpha \in \mathbb{C}$, we define

$$x^{\alpha} := \exp(\alpha \log(x)),$$

where $\exp, \log \in \mathbb{Q}[\![X]\!]$ are the standard formal power series corresponding to the exponential and the logarithmic maps.

Theorem 2.3. *Suppose that $\alpha \in \mathbb{C}$ and $[\mathbb{Q}(\alpha) : \mathbb{Q}] > n$. Let $x_1, \dots, x_n \in 1+X\mathbb{C}[\![X]\!]$ be multiplicatively independent. Then*

$$\operatorname{trdeg}_{\mathbb{C}(X)}(x_1, \dots, x_n, x_1^{\alpha}, \dots, x_n^{\alpha}) \geqslant n.$$

We now briefly describe modular analogues of Ax's theorem. Our disclaimer from the introduction applies very much here. Ax–Schanuel statements may go beyond the context of group homomorphisms: the first example here is the j-function map

$$j : \mathbb{H} \to \mathbb{C},$$

where $\mathbb{H}$ is the upper half plane. The linear independence assumption from Ax's theorem is replaced with *modular independence*. Pila's notes [2015] contain an excellent comprehensive survey of the state of the art in this field up to 2013. Such results have very important diophantine applications such as

- another proof of the Manin–Mumford conjecture [Pila and Zannier 2008];
- the first unconditional proof of the André–Oort conjecture for $\mathbb{C}^n$ [Pila 2011];
- a recent proof of the full André–Oort conjecture for Shimura varieties [Pila et al. 2021].

Following a suggestion by the referee, we would like to point out that only the Ax–Lindemann–Weierstrass type of results are needed in Manin–Mumford and André–Oort, while Ax–Schanuel (in fact, a weak form of it) is used in Zilber–Pink type problems.

In [Casale et al. 2020], the Ax–Lindemann–Weierstrass theorem with derivatives for the uniformizing functions of genus zero Fuchsian groups of the first kind is shown. This result is used in [Casale et al. 2020] to answer a question of Painlevé from 1895.

Remark 2.4. We analyze now which statements of the Ax–Schanuel results discussed above are transferable to the positive characteristic case. We would like to mention that all the analytic Ax–Schanuel type results over $\mathbb{C}$ may be replaced with their formal counterparts over an arbitrary field C, which was already done by Ax: the reader is advised to compare Ax's theorem on the dimension of intersections with its formal counterpart [Ax 1972, Theorem 3], which will be stated in a more general form in Section 3. Let us recall the setup first.

Definition 2.5 [Bochner 1946]. An n-dimensional *formal group* (law) over C is a tuple of power series $F \in C[\![X, Y]\!]^{\times n}$ ($|X| = |Y| = |Z| = n$) satisfying

$$F(0, X) = F(X, 0) = X,$$

$$F(X, F(Y, Z)) = F(F(X, Y), Z).$$

A *morphism* from an n-dimensional formal group G into an m-dimensional formal group F is a tuple of power series $f \in C[\![X]\!]^{\times m}$ such that

$$F(f(X), f(Y)) = f(G(X, Y)).$$

There is a well-known formalization functor $G \mapsto \widehat{G}$ (see pages 5 and 13 in [Manin 1963]) from the category of algebraic groups to the category of formal groups.

Such characteristic 0 formal statements seem to be transferrable to the positive characteristic context in the cases when the corresponding formal maps exist.

(1) The very original version of Ax–Schanuel does not look transferable, since there are no reasonable exponential maps in positive characteristic (we will briefly touch on the Drinfeld context at the end of Section 3).

(2) Therefore, other analytic maps need to be considered. "Analytic" may be replaced with "formal" (as mentioned above) and then the closest one to the exponential map which survives in the case of positive characteristic seems to be the formal isomorphism between the multiplicative group and an ordinary elliptic curve.

(3) The other types of such maps come from raising to powers in the multiplicative group.

Items (2) and (3) above will be discussed in the positive characteristic context in Section 3.

2B. ***Motivations and applications.*** Zilber [2002] used the differential Ax's theorem to prove *weak CIT*, which is a weak version of the *conjecture on intersection with tori* (CIT), which was also stated in [Zilber 2002]. CIT is a finiteness statement about intersections of subtori of a given torus with certain subvarieties of this torus. Weak CIT was used in [Baudisch et al. 2009] to produce a characteristic 0 *bad field*.

The existence of such a field was an open problem in model theory for almost 20 years.

Regarding the positive characteristic case, weak CIT does not hold and Zilber formulated a conjectural statement in (the very last statement of) [Zilber 2005]. It is still open whether a bad field in the positive characteristic case exists, however, Wagner [2003] showed that its existence in the case of characteristic $p > 0$ implies the existence of infinitely many *p-Mersenne primes*, which is an open problem in number theory — but it is widely believed that there are finitely many of them (for each individual prime p). Therefore, the existence of bad fields in positive characteristic looks very unlikely. However, pursuing the following path of research still looks interesting:

(1) prove positive characteristic versions of Ax–Schanuel;

(2) show a version of weak CIT in positive characteristic using (1);

(3) construct a version of a bad field in positive characteristic using (2);

(4) check the possible number-theoretic consequences of results obtained in (3).

As was mentioned in the previous subsection, Jonathan Pila and others used Ax–Schanuel type results to show different versions of the André–Oort conjecture; see, e.g., [Pila 2011; Tsimerman 2018; Casale et al. 2020; Pila et al. 2021].

There are also model-theoretic consequences of results of Ax–Schanuel type and we would like to point out some of them.

- Kirby [2009] used his version of an Ax–Schanuel statement to obtain the complete first-order theories of the exponential differential equations of semi-abelian varieties which arise from an amalgamation construction in the style of Hrushovski.
- Aslanyan [2022] did a version of the above for the j-function in place of the exponential function on semiabelian varieties.
- Freitag and Scanlon [2018] used Ax–Lindemann–Weierstrass to establish strong minimality and triviality of the differential equation of the j-function. This was generalized in [Aslanyan 2021] to a more general and formal setting.
- In [Casale et al. 2020] and [Blázquez-Sanz et al. 2021], the authors go in a quite opposite way: they first establish strong minimality using differential Galois theory, then use Zilber's trichotomy to prove triviality, then use that to establish Ax–Lindemann–Weierstrass and later Ax–Schanuel. That is, they give a new proof to Ax–Schanuel for the j-function and in fact for all Fuchsian automorphic functions.

3. Positive characteristic

The first (to my knowledge) positive characteristic Ax–Schanuel result concerns additive power series. Interestingly, it is not included in the cases considered in Remark 2.4, because such formal maps have no counterpart in the characteristic 0 case, since any additive formal power series in characteristic 0 is linear, so it is "not interesting". This positive characteristic additive Ax–Schanuel result is explained in detail below.

For any commutative algebraic group G, we have the following two (usually noncommutative) rings:

(1) the ring of *algebraic endomorphisms* (that is, endomorphisms of G in the original category of algebraic groups), denoted $\mathrm{End}_{\mathrm{algebraic}}(G)$;

(2) the ring of *formal endomorphisms* (that is, endomorphisms of the formalization of G, as below Definition 2.5, in the category of formal groups), denoted $\mathrm{End}_{\mathrm{formal}}(G)$.

Let C be a field of characteristic $p > 0$ and $\mathbb{G}_{\mathrm{a}}$ denote the additive group scheme over C. We consider the following two rings.

- The ring of additive polynomials over C (with composition), which we denote by $C[\mathrm{Fr}]$. This is also the skew polynomial ring over C and we have the ring isomorphism
$$\mathrm{End}_{\mathrm{algebraic}}(\mathbb{G}_{\mathrm{a}}) \cong C[\mathrm{Fr}].$$
- The ring of additive power series over C (with composition), which we denote by $C[\![\mathrm{Fr}]\!]$. We have the ring isomorphism
$$\mathrm{End}_{\mathrm{formal}}(\mathbb{G}_{\mathrm{a}}) \cong C[\![\mathrm{Fr}]\!].$$

These rings are commutative if and only if $C = \mathbb{F}_p$ and then they are also domains (isomorphic to the rings of polynomials or the ring of power series). We denote the fraction field of $\mathbb{F}_p[\mathrm{Fr}]$ by $\mathbb{F}_p(\mathrm{Fr})$. We state below the main theorem of [Kowalski 2012].

Ax–Schanuel for additive power series [Kowalski 2012, Theorem 1.1]. *Let F be an additive power series over $\mathbb{F}_p$ and assume that*
$$[\mathbb{F}_p(\mathrm{Fr})(F) : \mathbb{F}_p(\mathrm{Fr})] > n.$$
Let $x_1, \dots, x_n \in t\mathbb{F}_p[\![t]\!]$ be linearly independent over $\mathbb{F}_p[\mathrm{Fr}]$. *Then we have*
$$\mathrm{trdeg}_{\mathbb{F}_p}(x_1, \dots, x_n, F(x_1), \dots, F(x_n)) \geqslant n + 1.$$

We describe a general Ax–Schanuel result from [Kowalski 2019], which is valid in all characteristics. We need two technical assumptions. Before stating them, we try to motivate them. One of the crucial properties (used in the proofs in [Ax 1972]) of analytic homomorphisms between algebraic groups is that they take invariant algebraic differential forms into invariant algebraic differential forms. The first technical assumption below, which is absolutely necessary, is both formalizing and generalizing this crucial property. Regarding the second assumption, the exponential map gives a formal isomorphism between any commutative algebraic (and even formal) group in the case of characteristic 0 and a Cartesian power of the additive group. This is false in the positive characteristic case, for example there is no formal isomorphism between the additive and the multiplicative group (no exponential map in positive characteristic!). To make the proofs work, we still need to impose an additional assumption in the positive characteristic case, to mimic the above characteristic 0 situation. The 1-dimensional group H in this assumption plays the role of $\mathbb{G}_{\mathrm{a}}$ and we need to put some extra conditions on H which are true for $\mathbb{G}_{\mathrm{a}}$. We would prefer to avoid this second assumption, but we were unable to do so in [Kowalski 2019].

(1) We define a *special* formal map as one which "resembles a homomorphism" in the sense that it takes invariant differential forms into the "usual" differential forms (before taking the completion; see [Kowalski 2019, Definition 3.10]). In the positive characteristic case, the notion of differential forms has to be replaced by Vojta's notion [2007] of *higher differential forms*; see [Kowalski 2019, Remark 5.18(3)].

(2) We say that a commutative algebraic group A defined over the field C of characteristic p is "good" (see [Kowalski 2019, Definition 3.4]) if there is a one-dimensional algebraic group H over C such that we have the following (in the case of $p=0$, we drop (c)):
   (a) $\widehat{A} \cong \widehat{H^n}$.
   (b) The map $\mathrm{End}(\widehat{H}) \to \mathrm{End}_C(\Omega_H^{\mathrm{inv}})$ $(= C)$ is onto.
   (c) H is $\mathbb{F}_p$*-isotrivial*, i.e., $H \cong H^{\mathrm{Fr}}$.

To motivate the next result and give a general feeling regarding "what is it about", we quote from [Kowalski 2019] the following, where "the main theorem of this paper" refers to Theorem 3.1.

> A continuous map between Hausdorff spaces which is constant on a dense set is constant everywhere. The same principle applies to an algebraic map between algebraic varieties and to the Zariski topology (which is not Hausdorff). However, if we mix categories there is no reason for this principle to hold, e.g., there are nonconstant *analytic* maps between *algebraic* varieties which are constant on a Zariski dense subset. The main theorem of this paper roughly says that the principle above can be saved for certain formal maps (resembling homomorphisms) between

an algebraic variety and an algebraic group at the cost of replacing the range of the map with its quotient by a formal subgroup of the controlled dimension.

Theorem 3.1. *Let V be an algebraic variety, $\mathcal{K}$ a Zariski dense formal subvariety of V, A a "good" commutative algebraic group and $\mathcal{F} : \widehat{V} \to \widehat{A}$ a special formal map. Assume $\mathcal{F}$ vanishes on $\mathcal{K}$. Then there is a formal subgroup $\mathcal{C} \leqslant \widehat{A}$ such that $\mathcal{F}(\widehat{V}) \subseteq \mathcal{C}$ and*

$$\dim(\mathcal{C}) \leqslant \dim(V) - \dim(\mathcal{K}).$$

As a consequence of Theorem 3.1, we obtained in [Kowalski 2019] a result which is parallel to Ax–Schanuel for additive power series, where an additive power series (that is, a formal endomorphism of the additive group) is replaced with a "multiplicative" power series (that is, a formal endomorphism of the multiplicative group). Let $\mathbb{Z}_p$ denote the ring of p-adic integers. By [Hazewinkel 1978, Theorem 20.2.13(i)], we have the ring isomorphism

$$\mathrm{End}_{\mathrm{formal}}(\mathbb{G}_{\mathrm{m}}, \mathbb{G}_{\mathrm{m}}) \cong \mathbb{Z}_p.$$

We obtain an interesting positive characteristic version (see Example 4.15(3) and Theorem 4.16 in [Kowalski 2019]) of raising to powers Ax–Schanuel (see Theorem 2.3). For $x \in 1 + XC[\![X]\!]$ and $\alpha \in \mathbb{Z}_p$ ($\mathrm{char}(C) = p > 0$), we represent α as $\sum_{i=0}^{\infty} \alpha_i p^i$ for some $\alpha_i \in \{0, 1, \dots, p-1\}$ and we have

$$x^{\alpha} := \lim_{n\to\infty} \prod_{i=0}^{n} x^{\alpha_i p^i}.$$

Theorem 3.2. *Suppose that $\alpha \in \mathbb{Z}_p$ and $[\mathbb{Q}(\alpha):\mathbb{Q}] > n$. Let $x_1, \dots, x_n \in 1+XC[\![X]\!]$ be multiplicatively independent. Then*

$$\mathrm{trdeg}_{C(X)}(x_1, \dots, x_n, x_1^{\alpha}, \dots, x_n^{\alpha}) \geqslant n.$$

There is a general setup including the additive and multiplicative cases, which we describe below following [Kowalski 2019]. Let us fix a positive integer n and a one-dimensional algebraic group H over C. We introduce the following notation from [Kowalski 2019].

- Let $\boldsymbol{R} := \mathrm{End}_{\mathrm{algebraic}}(H)$ and $\boldsymbol{S} := \mathrm{End}_{\mathrm{formal}}(H)$.
- We restrict our attention to algebraic groups H such that $\boldsymbol{S}$ is a commutative domain. We regard $\boldsymbol{R}$ as a subring of $\boldsymbol{S}$.
- Let $\boldsymbol{K}$ denote the field of fractions of $\boldsymbol{R}$ and $\boldsymbol{L}$ be the field of fractions of $\boldsymbol{S}$. We regard $\boldsymbol{K}$ as a subfield of $\boldsymbol{L}$.

Example 3.3. In the characteristic 0 case, we always have $\boldsymbol{S} = C$, so the commutativity assumption is satisfied and we can consider any one-dimensional algebraic group as H. We give some examples below.

(1) If $H = \mathbb{G}_a$ and the characteristic is 0, then $\boldsymbol{R} = \boldsymbol{S} = C$.

(2) If $H = \mathbb{G}_a$ and the characteristic is $p > 0$, then $\boldsymbol{R} = C[\mathrm{Fr}]$ and $\boldsymbol{S} = C[\![\mathrm{Fr}]\!]$ (see the notation introduced in the beginning of this section). This is why we needed to take $C = \mathbb{F}_p$ to ensure that $\boldsymbol{S}$ is commutative.

(3) If $H = \mathbb{G}_m$ and the characteristic is 0, then $\boldsymbol{R} = \mathbb{Z}$. In the case of characteristic $p > 0$, we have $\boldsymbol{S} = \mathbb{Z}_p$ as mentioned above.

Below is our transcendental statement about formal endomorphisms (see [Kowalski 2019, Theorem 4.16.]). We need to introduce the following notions from [Kowalski 2019]. Let A be a commutative algebraic group over the field C of characteristic $p > 0$.

- A formal map into $\widehat{A}$ is an *A-limit map* if it can be "strongly approximated" by a sequence of polynomial maps $(f_n)_n$ in the sense that the differences $f_{n+1} - f_n$ are in the image of the n-th power of the appropriate Frobenius map. For example, any formal endomorphism of $\widehat{\mathbb{G}_a}$ is a $\mathbb{G}_a$-limit map (approximated by additive polynomials), and any formal endomorphism of $\widehat{\mathbb{G}_m}$ is a $\mathbb{G}_m$-limit map (approximated by multiplicative polynomials appearing in the description of x^α before the statement of Theorem 3.2).
- We fix a complete C-algebra $\mathcal{R}$ with the residue field C such that $\mathcal{R}$ is linearly disjoint from C^{alg} over C and in the case of characteristic p such that $L^{p^\infty} = C$, where L is the fraction field of $\mathcal{R}$ (e.g., $\mathcal{R}$ may be the power series algebra).
- For $x \in A(\mathcal{R})$, we call x *subgroup independent* if for any proper algebraic subgroup $A_0 < A$ defined over C, we have $x \notin A_0(\mathcal{R})$.
- The formal locus of $x \in A(\mathcal{R})$ over C is defined as the formal subscheme of $\widehat{A}$ corresponding to the image of the map $\widehat{\mathcal{O}}_{A,0} \to \mathcal{R}$.
- The number $\operatorname{andeg}(x)$ denotes the dimension of the formal locus of x over C.

Theorem 3.4. *Take $\gamma \in \boldsymbol{S}$ such that $[\boldsymbol{K}[\gamma] : \boldsymbol{K}] > n$ and $\gamma : \widehat{H} \to \widehat{H}$ is an H-limit map. Let $\mathcal{E} : \widehat{A} \to \widehat{A}$ be the n-th cartesian power of γ, where $A = H^n$. Then for any subgroup independent $x \in A(\mathcal{R})_*$ we have*

$$\operatorname{trdeg}_C(x, \mathcal{E}_K(x)) \geqslant n + \operatorname{andeg}_C(x).$$

We showed in [Kowalski 2019] that an unproved version of Theorem 3.1 without the "goodness" assumptions implies the following conjecture. This conjecture is important for the following reasons.

- If the field C has characteristic 0, then this conjecture is a theorem of Ax [1972, Theorem 1F].
- Ax [1972, Section 3] showed that in the case of characteristic 0 (Ax did not consider the positive characteristic case), [Ax 1972, Theorem 1F] implies the

Ax–Schanuel statements regarding the differential equation of the "appropriate" formal/analytic homomorphisms between algebraic groups (Ax focused on the exponential maps on semiabelian varieties). The corresponding implication holds in the positive characteristic case as well.

Main conjecture (arbitrary characteristic). *Let G be an algebraic group over a field C of arbitrary characteristic, $\widehat{G}$ the formalization of G at the origin and $\mathcal{A}$ a formal subgroup of $\widehat{G}$. Let $\mathcal{K}$ be a formal subscheme of $\mathcal{A}$ and let V be the Zariski closure of $\mathcal{K}$ in G. Then there is a formal subgroup $\mathcal{B}$ of $\widehat{G}$ which contains $\mathcal{A}$ and $\widehat{V}$ such that*

$$\dim(\mathcal{B}) \leqslant \dim(V) + \dim(\mathcal{A}) - \dim(\mathcal{K}).$$

We formulate below a specific statement which would follow from this main conjecture.

Specific conjecture. *Suppose that* $\operatorname{char}(C) = p > 0$ *and let* $\gamma : \widehat{\mathbb{G}_{\mathrm{m}}} \to \widehat{E}$ *be a formal isomorphism, where E is an ordinary elliptic curve. Let $x_1, \ldots, x_n \in 1 + XC[\![X]\!]$ be multiplicatively independent. Then*

$$\operatorname{trdeg}_{C(X)}(x_1, \ldots, x_n, \gamma(x_1), \ldots, \gamma(x_n)) \geqslant n.$$

This case seems to be related to the "interesting research paths (1)–(4)" from Section 2B. More precisely, the formal map appearing in the specific conjecture looks "closest" to the exponential map from the original Ax's theorem, which was used by Zilber to show weak CIT (see Section 2B).

We finish this section with a brief discussion of the case of the Drinfeld modules. Drinfeld [1974] introduced elliptic modules, which are now called *Drinfeld modules*. Drinfeld modules can be understood as certain homomorphisms between $\mathbb{F}_q[X]$ and $K[\mathrm{Fr}]$, where q is a power of p and $K = \mathbb{F}_q((\theta))$ is the non-Archimedean field of Laurent series over $\mathbb{F}_q$. An additive power series over K is associated to each Drinfeld module and this series is entire on K. A number of transcendence results for such additive power series was obtained; see, e.g., [Yu 1986]. To the best of my knowledge, such results never include a version of the full Ax–Schanuel statement. For a survey of this theory, we refer the reader to [Brownawell 1998]. Before the invention of Drinfeld modules, a special case of such a series was introduced by Carlitz, which is now called the *Carlitz exponential* and has the form

$$\exp_C = X + \sum_{i=1}^{\infty} \frac{X^{p^i}}{(\theta^{p^i} - \theta)(\theta^{p^i} - \theta^p)\cdots(\theta^{p^i} - \theta^{p^{i-1}})}.$$

There are several Schanuel type results for the Carlitz exponential (see [Denis 1995]) and a Carlitz exponential version of the (still open) conjecture on algebraic independence of logarithms of algebraic numbers was proved in [Papanikolas 2008, Theorem 1.2.6]. The power series we consider do not fit in the Drinfeld module framework, since we consider power series with constant coefficients, that is, there is no transcendental element θ in the coefficients of our additive power series.

4. Recent ideas and speculations

In this section, we describe some recent early stage developments concerning Ax–Schanuel type problems. One of them regards combining the results from [Blázquez-Sanz et al. 2021] with Ax's theorem on the dimension of intersections. The other one is about differential versions of Ax–Schanuel in positive characteristic.

4A. ***Towards a general statement of Ax–Schanuel.*** Ax–Schanuel statements for analytic/formal homomorphisms in the case of characteristic 0 have one "umbrella statement" from which they all follow, which is Ax's theorem on the dimension of intersections from Section 3. No such "umbrella statement" was known for Ax–Schanuel statements for the maps like the j-invariant map until the recent preprint [Blázquez-Sanz et al. 2021], where a general form of an Ax–Schanuel type result is given (see [Blázquez-Sanz et al. 2021, Theorem A]). In this statement, the algebraic group G is again back in the picture (e.g., $G = \mathrm{PGL}_2(\mathbb{C})$ in the case of the j-invariant map), but the statement is quite technical and it is phrased in terms of leaves of flat connections on G-principal bundles, where such a leaf plays the role of the analytic subgroup $\mathcal{A}$ from Ax's theorem on the dimension of intersections from Section 3.

Connection version of Ax–Schanuel [Blázquez-Sanz et al. 2021, Theorem A]. *Let ∇ be a G-principal flat connection on the algebraic bundle $P \to Y$ such that*

- *the algebraic group G is sparse*;
- *the Galois group of ∇ coincides with G.*

Let V be an algebraic subvariety of P and $\mathcal{L}$ be a horizontal leaf of ∇. If

$$\dim V < \dim(V \cap \mathcal{L}) + \dim G$$

then the projection of $V \cap \mathcal{L}$ in Y is contained in a ∇-special subvariety of Y.

Sparsity of the algebraic group G above means that there are no proper Zariski dense complex analytic subgroups of G. The notion of a "∇-special" is more technical; it is phrased in terms of the Galois group of a connection (see [Blázquez-Sanz et al. 2021, Definition 2.4]).

Unlike in the case of [Ax 1972, Theorem 1], no analytic subgroup appears in [Blázquez-Sanz et al. 2021, Theorem A], so this theorem does not generalize [Ax 1972, Theorem 1]. We propose such a generalization which encompasses both the connection version of Ax–Schanuel and [Ax 1972, Theorem 1]. It will appear in [Gogolok and Kowalski $\geq$ 2024].

Connection and subgroup Ax–Schanuel. *Let ∇ be a G-principal flat connection on the algebraic bundle $P \to Y$ such that the Galois group of ∇ coincides with G and*

- *V is an algebraic subvariety of P,*
- *$\mathcal{A}$ is an analytic subgroup of G,*
- *$\mathcal{L}$ is a horizontal leaf of ∇.*

Suppose that $\mathcal{V}$ is an analytic submanifold of $\mathcal{A}$ which is Zariski dense in V. If

$$\dim V < \dim(V \cap \mathcal{L}) + \dim G$$

then there is an analytic subgroup $\mathcal{H}$ of G such that

$$\dim \mathcal{H} < \dim(V) - \dim(\mathcal{V})$$

and $V \subseteq \mathcal{A}\mathcal{H}$.

The results mentioned above concern the case of characteristic 0. In the “main conjecture” from Section 3, the notion of “analytic” is replaced with the notion of “formal” (see Remark 2.4), which makes sense in the case of arbitrary characteristic. The connection version of Ax–Schanuel [Blázquez-Sanz et al. 2021, Theorem A] mentioned above has not been considered in the positive characteristic case before, since it requires an appropriate version of the notion of a connection in positive characteristic. This is work in progress [Gogolok and Kowalski $\geq$ 2024].

4B. ***Hasse–Schmidt differential Ax–Schanuel.*** Positive characteristic versions of the differential Ax’s theorem have not been studied yet. It is clear that we cannot consider the usual derivations anymore, since the constants of differential fields of positive characteristic contain the image of the Frobenius map, and hence there is no room for any transcendence. It looks natural in this case to replace the derivations with iterative Hasse–Schmidt derivations and the field of constants with the field of absolute constants. We give the necessary definitions below.

- A sequence $\partial = (\partial_n : R \to R)_{n\in\mathbb{N}}$ of additive maps on a ring is called an *HS-derivation* if ∂_0 is the identity map, and for all $n \in \mathbb{N}$ and $x, y \in R$, we have
$$\partial_n(xy) = \sum_{i+j=n} \partial_i(x)\partial_j(y).$$
- An HS-derivation ∂ is called *iterative* if for all $i, j \in \mathbb{N}$ we have
$$\partial_i \circ \partial_j = \binom{i+j}{i}\partial_{i+j}.$$

- If (K, ∂) is a field with a Hasse–Schmidt derivation, then its field of absolute constants is the intersection
$$\bigcap_{i=1}^{\infty} \ker(\partial_i).$$

The passages between the differential Ax–Schanuel and the power series Ax–Schanuel (described in Remark 2.1) work only one way for the positive characteristic case, since the power series ring has a natural iterative Hasse–Schmidt derivation on it. However, it is not clear how to proceed in the opposite way, so Hasse–Schmidt differential Ax–Schanuel type results need to be proved separately. This is work in progress [Gogolok and Kowalski ≥ 2024].

We state below two such results which will appear in [Gogolok and Kowalski ≥ 2024] to give a flavour of these kinds of Ax–Schanuel conditions. Assume that (K, ∂) is a field of characteristic $p > 0$ with a Hasse–Schmidt derivation and C is a field contained in the field of absolute constants of (K, ∂).

Additive Hasse–Schmidt differential Ax–Schanuel. *Let*
$$F = \sum_{m=0}^{\infty} c_m X^{p^m} \in \mathbb{F}_p[\![\mathrm{Fr}]\!]$$
and suppose that the algebraic degree of F over $\mathbb{F}_p(\mathrm{Fr})$ is greater than n. Take $x_1, \dots, x_n, y_1, \dots, y_n \in K$ such that $x_1, \dots, x_n$ are linearly independent over $\mathbb{F}_p[\mathrm{Fr}]$ and for all $i \in \{1, \dots, n\}$,
$$\begin{aligned} D_1(y_i - c_0 x_i) &= 0, \\ D_p(y_i - c_0 x_i - c_1 x_i^p) &= 0, \\ &\vdots \\ D_{p^m}(y_i - c_0 x_i - c_1 x_i^p - \cdots - c_m x^{p^m}) &= 0, \\ &\vdots \end{aligned}$$
Then we have
$$\mathrm{trdeg}_{\mathbb{F}_p}(x_1, y_1, \dots, x_n, y_n) \geqslant n + 1.$$

Multiplicative Hasse–Schmidt differential Ax–Schanuel. *Let*
$$\gamma = \sum c_i p^i \in \mathbb{Z}_p$$
and suppose that the algebraic degree of γ over $\mathbb{Q}$ is greater than n. Take $x_1, \dots, x_n, y_1, \dots, y_n \in K$ such that $x_1, \dots, x_n$ are multiplicatively independent and for all $i \in \{1, \dots, n\}$,

$$\begin{aligned} D_1(y_i x_i^{-c_0}) &= 0, \\ D_p(y_i x_i^{-c_0-c_1 p}) &= 0, \\ &\vdots \\ D_{p^m}(y_i x_i^{-c_0-c_1 p-\cdots-c_m p^m}) &= 0, \\ &\vdots \end{aligned}$$

Then we have

$$\operatorname{trdeg}_C(x_1, y_1, \ldots, x_n, y_n) \geqslant n+1.$$

References

[Aslanyan 2021] V. Aslanyan, "Ax–Schanuel and strong minimality for the j-function", *Ann. Pure Appl. Logic* **172**:1 (2021), art. id. 102871. MR Zbl

[Aslanyan 2022] V. Aslanyan, "Adequate predimension inequalities in differential fields", *Ann. Pure Appl. Logic* **173**:1 (2022), art. id. 103030. MR Zbl

[Ax 1971] J. Ax, "On Schanuel's conjectures", *Ann. of Math.* (2) **93** (1971), 252–268. MR Zbl

[Ax 1972] J. Ax, "Some topics in differential algebraic geometry, I: Analytic subgroups of algebraic groups", *Amer. J. Math.* **94** (1972), 1195–1204. MR Zbl

[Baudisch et al. 2009] A. Baudisch, M. Hils, A. Martin-Pizarro, and F. O. Wagner, "Die böse Farbe", *J. Inst. Math. Jussieu* **8**:3 (2009), 415–443. MR Zbl

[Bertrand 2008] D. Bertrand, "Schanuel's conjecture for non-isoconstant elliptic curves over function fields", pp. 41–62 in *Model theory with applications to algebra and analysis*, vol. 1, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **349**, Cambridge Univ. Press, 2008. MR Zbl

[Blázquez-Sanz et al. 2021] D. Blázquez-Sanz, G. Casale, J. Freitag, and J. Nagloo, "A differential approach to Ax–Schanuel, I", preprint, 2021. arXiv 2102.03384

[Bochner 1946] S. Bochner, "Formal Lie groups", *Ann. of Math.* (2) **47** (1946), 192–201. MR Zbl

[Brownawell 1998] W. D. Brownawell, "Transcendence in positive characteristic", pp. 317–332 in *Number theory* (Tiruchirapalli, India, 1996), edited by V. K. Murty and M. Waldschmidt, Contemp. Math. **210**, Amer. Math. Soc., Providence, RI, 1998. MR Zbl

[Brownawell and Kubota 1977] W. D. Brownawell and K. K. Kubota, "The algebraic independence of Weierstrass functions and some related numbers", *Acta Arith.* **33**:2 (1977), 111–149. MR Zbl

[Casale et al. 2020] G. Casale, J. Freitag, and J. Nagloo, "Ax–Lindemann–Weierstrass with derivatives and the genus 0 Fuchsian groups", *Ann. of Math.* (2) **192**:3 (2020), 721–765. MR Zbl

[Denis 1995] L. Denis, "Indépendance algébrique et exponentielle de Carlitz", *Acta Arith.* **69**:1 (1995), 75–89. MR Zbl

[Drinfeld 1974] V. G. Drinfeld, "Elliptic modules", *Mat. Sb.* (*N.S.*) **94(136)** (1974), 594–627. In Russian; translated in *Math. USSR-Sb.* **23**:4 (1976), 561–592. MR Zbl

[Freitag and Scanlon 2018] J. Freitag and T. Scanlon, "Strong minimality and the j-function", *J. Eur. Math. Soc.* (*JEMS*) **20**:1 (2018), 119–136. MR Zbl

[Gogolok and Kowalski ≥ 2024] J. Gogolok and P. Kowalski, "Connections and differential Ax–Schanuel in arbitrary characteristic", in preparation.

[Hazewinkel 1978] M. Hazewinkel, *Formal groups and applications*, Pure Appl. Math. **78**, Academic Press, New York, 1978. MR Zbl

[Kirby 2009] J. Kirby, "The theory of the exponential differential equations of semiabelian varieties", *Selecta Math.* (*N.S.*) **15**:3 (2009), 445–486. MR Zbl

[Kowalski 2008] P. Kowalski, "A note on a theorem of Ax", *Ann. Pure Appl. Logic* **156**:1 (2008), 96–109. MR Zbl

[Kowalski 2012] P. Kowalski, "Schanuel property for additive power series", *Israel J. Math.* **190** (2012), 349–363. MR Zbl

[Kowalski 2019] P. Kowalski, "Ax–Schanuel condition in arbitrary characteristic", *J. Inst. Math. Jussieu* **18**:6 (2019), 1157–1213. MR Zbl

[Lang 1966] S. Lang, *Introduction to transcendental numbers*, Addison-Wesley, Reading, MA, 1966. MR Zbl

[Manin 1963] J. I. Manin, "Theory of commutative formal groups over fields of finite characteristic", *Uspekhi Mat. Nauk* **18**:6(114) (1963), 3–90. In Russian; translated in *Russ. Math. Surv.* **18**:6 (1963), 1–83. MR Zbl

[Papanikolas 2008] M. A. Papanikolas, "Tannakian duality for Anderson–Drinfeld motives and algebraic independence of Carlitz logarithms", *Invent. Math.* **171**:1 (2008), 123–174. MR Zbl

[Pila 2011] J. Pila, "O-minimality and the André–Oort conjecture for $\mathbb{C}^n$", *Ann. of Math.* (2) **173**:3 (2011), 1779–1840. MR Zbl

[Pila 2015] J. Pila, "Functional transcendence via o-minimality", pp. 66–99 in *O-minimality and diophantine geometry*, edited by G. O. Jones and A. J. Wilkie, London Math. Soc. Lecture Note Ser. **421**, Cambridge Univ. Press, 2015. MR Zbl

[Pila and Tsimerman 2016] J. Pila and J. Tsimerman, "Ax–Schanuel for the j-function", *Duke Math. J.* **165**:13 (2016), 2587–2605. MR Zbl

[Pila and Zannier 2008] J. Pila and U. Zannier, "Rational points in periodic analytic sets and the Manin–Mumford conjecture", *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.* **19**:2 (2008), 149–162. MR Zbl

[Pila et al. 2021] J. Pila, A. N. Shankar, J. Tsimerman, H. Esnault, and M. Groechenig, "Canonical heights on Shimura varieties and the André–Oort conjecture", preprint, 2021. arXiv 2109.08788

[Seidenberg 1958] A. Seidenberg, "Abstract differential algebra and the analytic case", *Proc. Amer. Math. Soc.* **9** (1958), 159–164. MR Zbl

[Tsimerman 2018] J. Tsimerman, "The André–Oort conjecture for $\mathcal{A}_g$", *Ann. of Math.* (2) **187**:2 (2018), 379–390. MR Zbl

[Vojta 2007] P. Vojta, "Jets via Hasse–Schmidt derivations", pp. 335–361 in *Diophantine geometry* (Pisa, 2005), edited by U. Zannier, CRM Series **4**, Ed. Norm., Pisa, 2007. MR Zbl

[Wagner 2003] F. O. Wagner, "Bad fields in positive characteristic", *Bull. London Math. Soc.* **35**:4 (2003), 499–502. MR Zbl

[Yu 1986] J. Yu, "Transcendence and Drinfeld modules", *Invent. Math.* **83**:3 (1986), 507–517. MR Zbl

[Zilber 2002] B. Zilber, "Exponential sums equations and the Schanuel conjecture", *J. London Math. Soc.* (2) **65**:1 (2002), 27–44. MR Zbl

[Zilber 2005] B. Zilber, "Analytic and pseudo-analytic structures", pp. 392–408 in *Logic Colloquium 2000* (Paris, 2000), edited by R. Cori et al., Lect. Notes Log. **19**, Assoc. Symbol. Logic, Urbana, IL, 2005. MR Zbl

Received 17 Dec 2022. Revised 11 Oct 2023.

PIOTR KOWALSKI:

pkowa@math.uni.wroc.pl
Instytut Matematyczny, Uniwersytet Wrocławski, Wrocław, Poland

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.701

Analytic continuation and Zilber's quasiminimality conjecture

Alex J. Wilkie

In this article, which is dedicated to my friend and colleague Boris Zilber on the occasion of his 75th birthday, I put forward a strategy for proving his quasiminimality conjecture for the complex exponential field. That is, for showing that every subset of $\mathbb{C}$ definable in the expansion of the complex field by the complex exponential function is either countable or cocountable. In fact the strategy applies to any expansion of the complex field by a countable set of entire functions (in any number of variables) and is based on a certain property — an analytic continuation property — of the o-minimal structure obtained by expanding the ordered field of real numbers by the restrictions to compact boxes of the real and imaginary parts of the functions in the given set.

In a final section I discuss briefly the (rather limited) extent of our unconditional knowledge in the area.

After some reflection we, Boris and I, agreed that it was in July 1993 that he first asked me whether I had thought about the model theory of the complex exponential field. The occasion was Logic Colloquium '93, the European Summer Meeting of the Association for Symbolic Logic, which was held that year at Keele University in the UK. We were both invited to give plenary lectures. Boris spoke about his results on model-theoretic dimensions in complex (and ultrametric) analysis and I gave a talk on recent work with Angus (eventually published in the paper [Macintyre and Wilkie 1996]) concerning the decidability of the real exponential field, which was still my main concern. I certainly hadn't even considered the complex exponential. In fact the idea seemed absurd; the set of integers was definable and so as far as I was concerned it was not a tame structure. But Boris, then as now, had considerably more imagination, and deeper insights than I did into potentially good model-theoretic behaviour of familiar mathematical structures. If the set of reals was also definable then, he agreed, the situation would indeed be hopeless. But what if it wasn't? In fact, could it not be the case that every definable subset of $\mathbb{C}$ was either countable or cocountable? That remark had a profound effect on my mathematical

MSC2020: primary 03C64; secondary 03C65, 30E99.
Keywords: quasiminimality, analytic continuation, exponential fields.

life and was the motivational force for much of my research since that moment almost thirty years ago. It still is.

I soon realised that I could make progress on the problem — the *quasiminimality* problem as Boris called it — only through o-minimal theory and hard analysis: my knowledge of abstract stability theory and its generalizations was simply not strong enough, and certainly not up to a level to be able to follow and make a meaningful contribution to Boris's ingenious, beautiful and eventually highly successful construction of a "pseudoexponential" field using a combination of Hrushovski's predimensions and Shelah's theory of excellent classes. So my own approach started with the observation that the complex exponential function is definable in an o-minimal structure (when considered as a function from $\mathbb{R}^2$ to $\mathbb{R}^2$) provided that the imaginary part of its argument is restricted to a bounded interval. (One may take this structure to be, for example, the real ordered field expanded by both the real exponential function and by the sine function restricted to the interval $[0, \pi]$.) After several years I came up with about forty pages of notes purporting to contain, amongst other things, a proof that the complex exponential field was quasiminimal and I intended to present a sketch of the argument at an Oberwolfach meeting in July 2004.

Most readers of this article will know that Boris' conjecture is still unresolved and hence that there must have been a mistake, as indeed there was. But three things did come out of those notes. Firstly, I had established (correctly) a positive solution of Schanuel's conjecture for "generic" finite sequences of real numbers, and this rescued my Oberwolfach talk. (The result was eventually generalized and published in the paper [Bays et al. 2010] written jointly with two of the editors of this volume.) Secondly, in order to cope with the fact that the $2\pi i\mathbb{Z}$-periodicity of the complex exponential function is at complete variance with the whole ethos of o-minimality, I was forced to investigate whether, and if so how, points with integer coordinates could lie in sets defined by algebro-exponential equations, and thereby obtain a measure of "nontameness". I certainly did not want to assume the full version of Schanuel's conjecture (which would have settled this particular issue), but if it could be shown that either there were few such points or, if not, then at least one such point behaved generically, then the "generic Schanuel theorem" just mentioned could be used and this might be sufficient to control the periodicity. It didn't work but, and I apologize for a small personal digression at this point, the idea of quantitatively investigating the occurrence of points with integer coordinates in sets defined as above did appeal to me. In fact, why not do the same in the framework of general o-minimal structures?

So I temporarily abandoned work on the quasiminimality problem and initiated the project of counting integer points lying in sets definable in a fixed, but arbitrary, o-minimal structure. I had limited success at the time, only managing to prove a result in the case of one-dimensional sets. The paper [Wilkie 2004] was published in 2004 and came to the attention of Jonathan Pila who had been working on similar issues. He had obtained bounds of the same general nature as mine for rational points lying in low-dimensional, globally subanalytic sets (i.e., sets definable in the o-minimal structure $\mathbb{R}_{\text{an}}$), but was finding it difficult to put his arguments into a general setting that would smoothly facilitate an induction on dimension. Jonathan and I met soon after and agreed that the setting had to be o-minimality. Our point-counting theorem was published in [Pila and Wilkie 2006] and this is how, at least from my point of view, that result came about: my motivation for studying integer points in o-minimally definable sets was, apart from the fun of it, completely motivated by Boris' quasiminimality problem. I had no inkling of how, over the following fifteen years, the result would be applied, with huge success, by Jonathan and many others to problems in diophantine geometry.

Serendipity? Not exactly. For while Boris, as far as I know, did not foresee such applications, his conjectures, often fearless but always with sound intellectual bases, have been an inspirational source of research throughout our community even when, and perhaps especially when, that research takes unexpected directions.

I mentioned that three things came out of my 2004 notes. The third occupies the remainder of this paper. It concerns my strategy, which I still believe to be plausible, for proving the quasiminimality conjecture for the complex exponential field and, possibly, for many other expansions of the complex field by entire functions. I am very grateful to the editors for giving me this opportunity to explain the strategy despite the fact that, at the time of writing, it has not resulted in any definite results. The main theorems do establish the quasiminimality of a certain class of structures expanding the complex field but is conditional on a property of locally definable holomorphic functions, namely that they have analytic continuations along "generic" paths (i.e., those avoiding *obvious* singularities). The first theorem is precisely this, while the second provides a criterion for the continuation in purely complex analytic terms that avoid notions of general definability.

So let $\widetilde{\mathbb{R}}$ be a fixed o-minimal expansion of the real ordered field $\overline{\mathbb{R}} := \langle \mathbb{R}; <, +, \cdot \rangle$. Following [Peterzil and Starchenko 2001; 2003] we say that a complex valued function F of the n complex variables $z_1, \ldots, z_n$ is *definable* if the real and imaginary parts of F are definable in $\widetilde{\mathbb{R}}$ (*without* parameters) when considered as functions of the $2n$ real variables $x_1, y_1, \ldots, x_n, y_n$, where $z_j = x_j + i y_j$ for $j = 1, \ldots, n$. If F is holomorphic, its domain will always be assumed to be, without further mention, a connected, open subset (a fortiori a *definable* subset if F is definable) of $\mathbb{C}^n$. We write $\operatorname{dom}(F)$ for the domain of F.

Peterzil and Starchenko developed complex analysis in this definable context, but in fact many of the subtleties of their work will not be needed here since we will only be working with the standard structure $\widetilde{\mathbb{R}}$. However, one of their results is still worth mentioning, namely that if $\{F_t : t \in \mathbb{R}^d\}$ is a definable family of n-variable, holomorphic functions, then there exists N such that for all $t \in \mathbb{R}^d$ and all $a \in \mathrm{dom}(F_t)$, if $\partial^\alpha F/\partial z^\alpha$ vanishes at a for each $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq N$, then F_t is identically zero. (This is, of course, obvious if $\widetilde{\mathbb{R}}$ is polynomially bounded, but not so clear otherwise. For example, and of particular relevance to Zilber's problem, the F_t's could range over polynomials in $z_1, \dots, z_n, \log z_1, \dots, \log z_n$ of some fixed degree (and with suitable simply connected domains).)

Anyway, returning to definitions, we consider the closure operator $\mathrm{LD}(\cdot)$ — "locally definable from" — where we specify that for all $n \geq 1$ and all $a_1, \dots, a_{n+1} \in \mathbb{C}$,

$$a_{n+1} \in \mathrm{LD}(a_1, \dots, a_n) \iff F(a_1, \dots, a_n) = a_{n+1} \tag{1}$$

for some definable, holomorphic function F with $\langle a_1, \dots, a_n\rangle \in \mathrm{dom}(F)$.

It is straightforward to show that LD is a pregeometry and, further, that for all $b_1, \dots, b_m \in \mathbb{C}$,

> $b_1, \dots, b_m$ are LD-independent if and only if whenever G is a nonzero, definable, holomorphic function with $\langle b_1, \dots, b_m\rangle \in \mathrm{dom}(G)$, then $G(b_1, \dots, b_m) \neq 0$. (2)

(See [Wilkie 2008] for details.)

Strictly speaking $\mathrm{LD}(X)$ should be specified for all subsets X of $\mathbb{C}$ and this is done in the usual way by taking $\mathrm{LD}(X)$ to be the union of all $\mathrm{LD}(X_0)$'s as X_0 ranges over finite subsets of X. There is a small issue concerning the LD-closure of the empty set, which is taken to be $\{s+it : s, t \in \mathbb{R}, \text{ both definable in } \widetilde{\mathbb{R}}\}$. One easily checks (using the Cauchy–Riemann equations) that if G is a nonzero, definable, holomorphic function with $s+it \in \mathrm{dom}(G)$ and $G(s+it) = 0$, then both s and t are indeed definable in $\widetilde{\mathbb{R}}$, so this choice of $\mathrm{LD}(\varnothing)$ is consistent with (2) above. However, the relationship between the LD-dimension of an arbitrary finite subset of $\mathbb{C}$ and the dimension of the corresponding set of real and imaginary parts (for the usual pregeometry of the o-minimal structure $\widetilde{\mathbb{R}}$) is more complicated and is resolved in Section 4 of [Wilkie 2008].

Examples. (a) If we take $\widetilde{\mathbb{R}}$ to be just $\overline{\mathbb{R}}$, then LD is just algebraic closure (over $\mathbb{Q}$) in the field $\mathbb{C}$. More generally, if we expand $\overline{\mathbb{R}}$ by a set A of constants, then LD is algebraic closure over the subfield $\mathbb{Q}(A)$ of $\mathbb{C}$. (Two remarks: firstly, I do not find this at all obvious (the proof can be found in [Wilkie 2005]), and secondly, recall that our holomorphic functions are allowed to be definable in $\widetilde{\mathbb{R}}$ so, for example, we can distinguish (uniformly in parameters) between different roots of a polynomial.)

(b) Let $\widetilde{\mathbb{R}} = \mathbb{R}_{\exp} := \langle\overline{\mathbb{R}}, \exp\rangle$, where $\exp : \mathbb{R} \to \mathbb{R} : x \mapsto e^x$. Then, by a result of Bianconi [2005], LD is still just algebraic closure (but this time it's over the minimal model $K \preccurlyeq \mathbb{R}_{\exp}$ of the theory of $\mathbb{R}_{\exp}$). Of course, things are different, and of great relevance to Zilber's problem, if we expand $\mathbb{R}_{\exp}$ by the restricted sine function and we will be discussing this situation later.

(c) For $\widetilde{\mathbb{R}} = \mathbb{R}_{\mathrm{an}}$ (where the definable sets are the globally subanalytic sets — see, for example, [Denef and van den Dries 1988]), LD is trivial: $\mathrm{LD}(X) = \mathrm{LD}(\varnothing) = \mathbb{C}$ for all $X \subseteq \mathbb{C}$ simply because all $r \in \mathbb{R}$ are definable.

In view of example (c) above we now assume that the language of $\widetilde{\mathbb{R}}$ is countable so, in particular, $\mathrm{LD}(\varnothing)$ is countable. We shall require rather more:

Lemma (existence of generic lines). *Let A be a countable subset of $\mathbb{R}$. Then there exists $\omega \in \mathbb{C}$ (in fact an uncountable, dense set of them) such that for all continuous functions $\phi : [0, 1] \mapsto \mathbb{C}$ definable in the structure $\langle \widetilde{\mathbb{R}}, s\rangle_{s\in A}$ and satisfying $\phi(0) \neq 0$, we have that $\phi(t) \neq t\omega$ for all $t \in [0, 1]$.*

Proof. For each such ϕ let

$$S_\phi := \{\omega \in \mathbb{C} : \text{for all } t \in [0, 1], \phi(t) \neq t\omega\}.$$

It clearly follows from the continuity of ϕ and the compactness of the closed unit interval that S_ϕ is an open subset of $\mathbb{C}$. So by the Baire category theorem we will be done if we show that S_ϕ is a dense subset of $\mathbb{C}$ (because there are only countably many ϕ's).

So let ω_0 be an arbitrary complex number and let $\epsilon > 0$. Let Δ be the open disc in $\mathbb{C}$ centred at ω_0 and of radius ϵ and suppose, for a contradiction, that $\Delta \subseteq \mathbb{C} \setminus S_\phi$, i.e., that for all $\omega \in \Delta$, there exists $t \in [0, 1]$ such that $\phi(t) = t\omega$. Then, by definable choice and since dimension is nondecreasing under definable, injective maps (working in the o-minimal structure $\langle \widetilde{\mathbb{R}}, s\rangle_{s\in A}$), there exist $\omega_1, \omega_2 \in \Delta$ with $\omega_1 \neq \omega_2$ such that for some $t_0 \in [0, 1]$, $\phi(t_0) = t_0\omega_1$ and $\phi(t_0) = t_0\omega_2$. This is absurd unless $t_0 = 0$; but this is also ruled out since $\phi(0) \neq 0$. □

We will be considering analytic continuations of definable functions along generic paths in $\mathbb{C}^n$. In fact, we only need to consider linear paths: for $a, \omega \in \mathbb{C}^n$ $(n \geq 1)$, define the map $\lambda_{a,\omega} : \mathbb{C} \mapsto \mathbb{C}^n$ by $\lambda_{a,\omega}(z) := a + z\omega$ (for $z \in \mathbb{C}$). We say that $\lambda_{a,\omega}$ is *generic* on a set $T \subseteq \mathbb{C}$ if $\lambda_{a,\omega}(t)$ is a generic n-tuple for each $t \in T$, i.e., if $a_1 + t\omega_1, \ldots, a_n + t\omega_n$ are LD-independent complex numbers for each $t \in T$ (where $a = \langle a_1, \ldots, a_n\rangle$ and $\omega = \langle \omega_1, \ldots, \omega_n\rangle$). The set T is almost always the interval $[0, 1]$, so that for any a, ω we have that $\lambda_{a,\omega}(0) = a$.

We now come to our main definitions.

Definition 1. We say that the structure $\widetilde{\mathbb{R}}$ has the *analytic continuation property* (ACP) if for all LD-independent $a_1, \ldots, a_n \in \mathbb{C}$, all definable, holomorphic functions F with $a = \langle a_1, \ldots, a_n\rangle \in \mathrm{dom}(F)$ and all $\omega \in \mathbb{C}^n$ with $\lambda_{a,\omega}$ generic on $[0, 1]$, there exists a definable, holomorphic function G with $\lambda_{a,\omega}([0, 1]) \subseteq \mathrm{dom}(G) \subseteq \mathbb{C}^n$ such that $G(a) = F(a)$. (And hence, by (2), $G(\lambda_{a,\omega}(z)) = F(\lambda_{a,\omega}(z))$ for all $z \in \mathbb{C}$ such that $\lambda_{a,\omega}(z)$ lies in the connected component of $\mathrm{dom}(F) \cap \mathrm{dom}(G)$ containing the point a. So the function $G \circ \lambda_{a,\omega}$ analytically continues, in the usual sense, the function $F \circ \lambda_{a,\omega}$ (restricted to a sufficiently small open neighbourhood of $0 \in \mathbb{C}$) to an open set containing the interval [0, 1].)

Definition 2. Let $1 \le l \le n$ and let $M \subseteq \mathbb{C}^n$. Then we say that M is an *l-dimensional*, *locally definable*, *complex submanifold of* $\mathbb{C}^n$ (or just an *l-manifold* for short) if

(a) M is a closed subset of $\mathbb{C}^n$, and

(b) for all $a \in M$, there exist a definable open set W with $a \in W \subseteq \mathbb{C}^n$ and a holomorphic, definable map $G = \langle G_1, \dots, G_{n-l} \rangle : W \to \mathbb{C}^{n-l}$ such that a is a nonsingular point of the zero set of G (i.e., $G(a) = 0$ and the vectors $\langle (\partial G_j / \partial z_1)(a), \dots, (\partial G_j / \partial z_n)(a) \rangle$ (for $1 \le j \le n-l$) are linearly independent over $\mathbb{C}$) and, further, $M \cap W = Z_{\mathrm{reg}}(G)$, where $Z_{\mathrm{reg}}(G)$ denotes the set of nonsingular points of the zero set of G.

Example. If $\widetilde{\mathbb{R}} = \langle \overline{\mathbb{R}}, \exp \restriction [0, 1], \sin \restriction [0, 2\pi] \rangle$, then the graph of the complex exponential function $\Gamma_{\exp} := \{\langle z, e^z \rangle : z \in \mathbb{C}\}$ is a 1-dimensional, locally definable, complex submanifold of $\mathbb{C}^2$.

Definition 3. The structure $\widetilde{\mathbb{C}}$ is defined to be the expansion of the complex field by all l-dimensional, locally definable, complex submanifolds of $\mathbb{C}^n$ (for all l, n with $1 \le l < n$).

Remark. With $\widetilde{\mathbb{R}}$ as in the example above, we see that $\widetilde{\mathbb{C}}$ is an expansion of $\mathbb{C}_{\exp}$ (:= the complex field expanded by the complex exponential function). I do not know if it is a proper expansion. For example, it is clear that every connected component of an l-manifold is also an l-manifold, but it seems to me (and, in fact, to Zilber) to be perfectly possible that some l-manifold, M say, is definable in $\mathbb{C}_{\exp}$ but that some connected component of M is not.

Theorem 4. *Suppose that $\widetilde{\mathbb{R}}$ has the ACP. Then the structure $\widetilde{\mathbb{C}}$ is quasiminimal. That is, for any subset S of $\mathbb{C}$ which is definable in the language $\mathcal{L}(\widetilde{\mathbb{C}})$ of $\widetilde{\mathbb{C}}$ (and we do allow parameters here), we have that either S is countable or its complement $\mathbb{C} \setminus S$ is countable. In fact, the same is true for sets S defined by a formula of the infinitary language $\mathcal{L}(\widetilde{\mathbb{C}})_{\infty,\omega}$ provided, of course, that the formula contains only countably many parameters.*

Proof. We first show that if u and v are elements of $\mathbb{C} \setminus \mathrm{LD}(\varnothing)$ then there exists a back-and-forth system (for the structure $\widetilde{\mathbb{C}}$) containing the pair $\langle u, v \rangle$.

Then by a classical result of Karp [1964] the quasiminimality condition for parameter-free formulas follows from this since it implies that if $S \subseteq \mathbb{C}$ is a parameter-free $\mathcal{L}(\widetilde{\mathbb{C}})_{\infty,\omega}$-definable set, then either $S \subseteq \mathrm{LD}(\varnothing)$ or else $\mathbb{C} \setminus \mathrm{LD}(\varnothing) \subseteq S$, and $\mathrm{LD}(\varnothing)$ is countable.

So suppose that $u, v \in \mathbb{C}\setminus\mathrm{LD}(\varnothing)$. We may assume that $\lambda_{u,v-u}$ is generic on $[0, 1]$. For otherwise, by the countability of $\mathrm{LD}(\varnothing)$, there exists some $w \in \mathbb{C}\setminus\mathrm{LD}(\varnothing)$ such that both $\lambda_{u,w-u}$ and $\lambda_{w,v-w}$ are generic on $[0, 1]$ and we prove the result for the pair u, w and for the pair w, v.

We now set up a back-and-forth argument.

For $n \geq 1$ and $a = \langle a_1, \dots, a_n\rangle$, $b = \langle b_1, \dots, b_n\rangle \in \mathbb{C}^n$ we write $a \sim_n b$ if

(i)$_n$ $a_1 = u$ and $b_1 = v$;

(ii)$_n$ for some m with $1 \leq m \leq n$ and some $i_1, \dots, i_m$ with $1 \leq i_1 < \cdots < i_m \leq n$ we have that $\lambda_{a',b'-a'}$ is generic on $[0, 1]$, where $a' := \langle a_{i_1}, \dots, a_{i_m}\rangle$ and $b' := \langle b_{i_1}, \dots, b_{i_m}\rangle$;

(iii)$_n$ there exists a definable, connected, open set $V \subseteq \mathbb{C}^m$ with $\lambda_{a',b'-a'}([0, 1]) \subseteq V$ and, for each $j = 1, \dots, n$, a definable, holomorphic function $F_j : V \to \mathbb{C}$ such that $F_j(\lambda_{a',b'-a'}(0)) = F_j(a_{i_1}, \dots, a_{i_m}) = a_j$ and $F_j(\lambda_{a',b'-a'}(1)) = F_j(b_{i_1}, \dots, b_{i_m}) = b_j$.

We clearly have $\langle u\rangle \sim_1 \langle v\rangle$ (where, in this case, $n = m = 1$, $V = \mathbb{C}$ and F_1 is the identity function on $\mathbb{C}$).

In order to establish the back-and-forth property, suppose that $n \geq 1$ and that $a \sim_n b$ as above, with m, a', b' as in (ii)$_n$. We write λ for $\lambda_{a',b'-a'}$.

So let $a_{n+1} \in \mathbb{C}$. We must find $b_{n+1} \in \mathbb{C}$ such that $\langle a, a_{n+1}\rangle \sim_{n+1} \langle b, b_{n+1}\rangle$. There are two cases.

Case 1. $a_{n+1} \notin \mathrm{LD}(a_{i_1}, \dots, a_{i_m})$.

Let A be a finite subset of $\mathbb{R}$ containing the real and imaginary parts of $a_{i_1}, \dots, a_{i_m}$, $a_{n+1}, b_{i_1}, \dots, b_{i_m}$. Apply the generic lines lemma to obtain some $\omega \in \mathbb{C}$ such that for all continuous $\phi : [0, 1] \to \mathbb{C}$ definable in the structure $\langle \widetilde{\mathbb{R}}, c\rangle_{c\in A}$ with $\phi(0) \neq 0$, we have $\phi(t) \neq t\omega$ for all $t \in [0, 1]$ and, further, such that ω does not lie in the (countable) set $\mathrm{LD}(b_{i_1}, \dots, b_{i_m}, a_{n+1})$. Let $b_{n+1} := a_{n+1} + \omega$. We show that $\langle a, a_{n+1}\rangle \sim \langle b, b_{n+1}\rangle$. Now (i)$_{n+1}$ is obvious and for (ii)$_{n+1}$, (iii)$_{n+1}$ we replace m by $m + 1$ ($\leq n + 1$) and let $i_{m+1} := n + 1$. We also replace a' by $\langle a', a_{n+1}\rangle$ and b' by $\langle b', b_{n+1}\rangle$ so that λ ($= \lambda_{a',b'-a'}$) becomes $\lambda^* : \mathbb{C} \to \mathbb{C}^{m+1}$ given by $\lambda^*(z) := \langle \lambda(z), a_{n+1} + z\omega\rangle$. We must show that λ^* is generic on $[0, 1]$. So suppose, for a contradiction, that for some $t_0 \in [0, 1]$ the $(m + 1)$-tuple $\lambda^*(t_0)$ has LD-dimension $< m + 1$. Now since $\lambda(t_0)$ has LD-dimension m (because λ is generic on $[0, 1]$) we must have $a_{n+1} + t_0\omega \in \mathrm{LD}(\lambda(t_0))$. So there exists a definable, holomorphic function F with $\lambda(t_0) \in \mathrm{dom}(F)$ and such that

$$F(\lambda(t_0)) = a_{n+1} + t_0\omega. \tag{$*$}$$

Now we cannot have $t_0 = 0$, for then we would have $a' = \lambda(0) \in \mathrm{dom}(F)$ and $F(a') = a_{n+1}$, which contradicts the hypothesis of Case 1. Also, $t_0 \neq 1$, for otherwise we would have that $\omega = F(\lambda(1)) - a_{n+1} = F(b') - a_{n+1} \in \mathrm{LD}(b_{i_1}, \dots, b_{i_m}, a_{n+1})$. So $0 < t_0 < 1$ and since $\mathrm{dom}(F)$ is open and λ is continuous, there exist rationals q_1, q_2 with $0 < q_1 < t_0 < q_2 < 1$ such that $\lambda([q_1, q_2]) \subseteq \mathrm{dom}(F)$.

Define $\phi : [0, 1] \to \mathbb{C}$ to be the continuous function which takes the value $-a_{n+1} + F(\lambda(t))$ for $q_1 \le t \le q_2$ and which is linear on the intervals $[0, q_1]$, $[q_2, 1]$ with, say, $\phi(0) = 1$ and $\phi(1) = 0$.

Then ϕ is definable in the structure $\langle \widetilde{\mathbb{R}}, c\rangle_{c\in A}$ because λ is, a_{n+1} is, and F is definable in $\widetilde{\mathbb{R}}$ without parameters. Further, $\phi(0) \neq 0$ so by the construction of ω we have that $\phi(t) \neq t\omega$ for all $t \in [0, 1]$. In particular, $\phi(t_0) \neq t_0$, i.e., $-a_{n+1} + F(\lambda(t_0)) \neq t_0\omega$, which contradicts $(*)$ and establishes $(\mathrm{ii})_{n+1}$.

As for $(\mathrm{iii})_{n+1}$, we take our new V to be $V \times \mathbb{C}$ and our new F_j's — call them F_j^* for $j = 1, \dots, n+1$ — to be specified (for $\langle z_1, \dots, z_{m+1}\rangle \in V \times \mathbb{C}$) by setting $F_j^*(z_1, \dots, z_{m+1}) := F_j(z_1, \dots, z_m)$ if $1 \le j \le n$ and $F_{n+1}^*(z_1, \dots, z_{m+1}) := z_{m+1}$.

Then the required conditions for $(\mathrm{iii})_{n+1}$ carry over from $(\mathrm{iii})_n$ for $j = 1, \dots, n$, and for $j = n+1$ we have, for each $t \in [0, 1]$,

$$F_{n+1}^*(\lambda_{\langle a', a_{n+1}\rangle, \langle b', b_{n+1}\rangle - \langle a', a_{n+1}\rangle}(t)) = a_{n+1} + t(b_{n+1} - a_{n+1}),$$

which takes the value a_{n+1} for $t = 0$ and b_{n+1} for $t = 1$. So we have that $\langle a, a_{n+1}\rangle \sim_{n+1} \langle b, b_{n+1}\rangle$ as required.

Case 2. $a_{n+1} \in \mathrm{LD}(a_{i_1}, \dots, a_{i_m})$.

In this case there is a definable, holomorphic function F such that $a' \in \mathrm{dom}(F)$ and $F(a') = a_{n+1}$. Apply the ACP to obtain a definable, holomorphic function G with $\lambda([0, 1]) \subseteq \mathrm{dom}(G) \subseteq \mathbb{C}^m$ satisfying $G(a') = F(a')$, i.e., $G(a') = a_{n+1}$. Now with $V \subseteq \mathbb{C}^m$ as given by $(\mathrm{iii})_n$, note that $\lambda([0, 1]) \subseteq V \cap \mathrm{dom}(G)$. Let U be a definable, connected open subset of $V \cap \mathrm{dom}\, G$ such that $\lambda([0, 1]) \subseteq U$. (Clearly such a U exists and may be taken, for example, to be a certain finite union of polydiscs with Gaussian rational centres and rational radii.)

We now take the same m as in $(\mathrm{ii})_n$ so that $(\mathrm{i})_{n+1}$ and $(\mathrm{ii})_{n+1}$ are obviously satisfied. For $(\mathrm{iii})_{n+1}$ we take the F_j's as given by $(\mathrm{iii})_n$ for $j = 1, \dots, n$ and restrict them to the set U. For F_{n+1} we take G restricted to U so that $F_{n+1}(\lambda(0)) = F_{n+1}(a') = G(a') = a_{n+1}$. Finally, taking $b_{n+1} := F_{n+1}(\lambda(1))$ completes the construction in Case 2.

Of course we also need to consider the "back" case, where we take some $b_{n+1} \in \mathbb{C}$ and have to find some $a_{n+1} \in \mathbb{C}$ satisfying $\langle a, a_{n+1}\rangle \sim \langle b, b_{n+1}\rangle$. But this follows in exactly the same way upon noting that for any $c, d \in \mathbb{C}^l$ we have that $\lambda_{c,d-c}$ is generic on $[0, 1]$ if and only if $\lambda_{d,c-d}$ is (because the ranges on $[0, 1]$ are the same).

So, our system $\{\langle a, b\rangle : n \ge 1, a, b \in \mathbb{C}^n \text{ and } a \sim_n b\}$ has the back-and-forth property.

We must now show that atomic formulas of $\mathcal{L}(\widetilde{\mathbb{C}})$ are preserved from a to b whenever $a \sim_n b$.

So let $1 \le l \le n$ and suppose that $M \subseteq \mathbb{C}^n$ is an l-manifold. Suppose $a \in M$ and $b \in \mathbb{C}^n$ are such that $a \sim_n b$. We must show that $b \in M$ (and similarly for a and b interchanged, for which the proof is the same). Write $a = \langle a_1, \ldots, a_n \rangle$ and $b = \langle b_1, \ldots, b_n \rangle$. Let m, $1 \le i_1 < \cdots < i_m \le n$, V and the F_j's be as in (ii)$_n$ and (iii)$_n$, and write F for the map $\langle F_1, \ldots, F_n \rangle : V \to \mathbb{C}^n$.

Define $T := \{t \in [0, 1] : F(\lambda(t)) \in M\}$, where, as before, $\lambda = \lambda_{a',b'-a'}$ and $a' = \langle a_{i_1}, \ldots, a_{i_m} \rangle$, $b' = \langle b_{i_1}, \ldots, b_{i_m} \rangle$,

Then T is not empty because $0 \in T$. Also, T is a closed subset of $[0, 1]$ because M is closed (see Definition 2(a)). So we shall be done if we can show that T is open.

So let $t_0 \in [0, 1]$ be such that $F(\lambda(t_0)) \in M$. Say $F(\lambda(t_0)) = c = \langle c_1, \ldots, c_n \rangle$. Choose $G = \langle G_1, \ldots, G_{n-l} \rangle$ and W as in Definition 2(b) for this particular $c \in M$. Then $c \in Z_{\mathrm{reg}}(G)$, and by reducing W (definably) if necessary we may suppose that w is a nonsingular point of the zero set of G for *all* $w \in W$ satisfying $G(w) = 0$. By continuity, there exists $\epsilon > 0$ such that $F(\lambda(t)) \in W$ for all $t \in [t_0 - \epsilon, t_0 + \epsilon]$. Thus $\{z \in V : F(z) \in W\}$ is a definable, open subset of V containing $\lambda([t_0 - \epsilon, t_0 + \epsilon])$ and $G \circ F : \{z \in V : F(z) \in W\}^* \to \mathbb{C}$ is a definable holomorphic function such that $G \circ F(\lambda(t_0)) = 0$, where the $*$ denotes taking the connected component of the set $\{z \in V : F(z) \in W\}$ that contains the point $\lambda(t_0)$ (and hence the set $\lambda([t_0 - \epsilon, t_0 + \epsilon])$). However, λ is generic on $[0, 1]$ and so it follows from (2) that $G \circ F$ is identically zero. In particular, $G(F(\lambda(t))) = 0$ for all $t \in [t_0 - \epsilon, t_0 + \epsilon]$. But $Z_{\mathrm{reg}}(G) \subseteq M \cap W$ (Definition 2(b)) and so $F(\lambda(t)) \in M$ for all $t \in [t_0 - \epsilon, t_0 + \epsilon]$, and this shows that T is open, as required.

The proof of our present aim is now complete apart from one small detail. The reader may have noticed that, strictly speaking, atomic formulas of $\mathcal{L}(\widetilde{\mathbb{C}})$ have the form $\Phi(v_{j_1}, \ldots, v_{j_p})$ for some $1 \le l < p$ where Φ is the symbol of the language $\mathcal{L}(\mathbb{C})$ interpreting an l-submanifold of $\mathbb{C}^p$ (for some l, p with $1 \le l < p$). But we have tacitly assumed in our proof above that $j_k = k$ and that p is (an arbitrarily large) n. But this assumption can easily be arranged (at least for $j_1, \ldots, j_p$ distinct) by "adding vacuous variables" and observing that the set

$$\{\langle a_1, \ldots, a_n \rangle \in \mathbb{C}^n : \widetilde{\mathbb{C}} \models \Phi[a_{j_1}, \ldots, a_{j_p}]\}$$

is an $(n - p + l)$-submanifold of $\mathbb{C}^n$.

Notice also that the graph of equality is a 1-submanifold of $\mathbb{C}^2$ (so we may indeed assume in the above discussion that $j_1, \ldots, j_p$ are distinct) and that the graphs of addition and multiplication are 2-submanifolds of $\mathbb{C}^3$, so that equality of polynomial terms is also preserved by the $\sim_n$ relation.

We now need to deal with the case that the formula defining the set S contains a countable set, X say, of parameters. But for this we simply apply the result above with $\widetilde{\mathbb{R}}$ replaced by the structure, $\breve{\mathbb{R}}$ say, obtained by expanding $\widetilde{\mathbb{R}}$ by a constant for each element of X', where X' is the set of real and imaginary parts of elements of X. The required result follows since it is easy to check (upon denoting by $\breve{\mathbb{C}}$ the corresponding complex structure as given by Definition 3) that for any formula of $\mathcal{L}(\widetilde{\mathbb{C}})_{\infty,\omega}$ with parameters in X there exists a parameter-free formula of $\mathcal{L}(\breve{\mathbb{C}})_{\infty,\omega}$ that defines the same set. $\square$

As a test for quasiminimality Theorem 4 has limited use because in order to prove that a given structure has the ACP one still needs some reasonable mathematical description of the definable sets and functions. We now turn to this problem in the case of expansions of the complex field by entire functions and we look for a *complex analytic* criterion for such a structure to have the ACP. To this end, suppose that we are given, for each $n \geq 0$, a countable ring $\mathcal{H}_n$ of entire functions of the n complex variables $z_1, \ldots, z_n$. We assume that $\mathcal{H}_n \subseteq \mathcal{H}_{n+1}$ (in the obvious sense) and that each $\mathcal{H}_n$ contains the projection functions and is closed under partial differentiation and Schwarz reflection (i.e., if $f \in \mathcal{H}_n$, then $\partial f/\partial z_j \in \mathcal{H}_n$ for $j = 1, \ldots, n$ and $f^{\mathrm{SR}} \in \mathcal{H}_n$, where $f^{\mathrm{SR}}(z) := \overline{f(\bar{z})}$ for $z \in \mathbb{C}^n$ (and the bar denotes coordinatewise complex conjugation)). We then call the sequence $\mathcal{H} := \langle \mathcal{H}_n : n \geq 0 \rangle$ of rings a *suitable sequence* and associate to such an $\mathcal{H}$ a certain expansion $\widetilde{\mathbb{R}}(\mathcal{H})$ of $\overline{\mathbb{R}}$ as follows:

> For each $n \geq 0$, $f \in \mathcal{H}_n$ and discs $D_1, \ldots, D_n$ in $\mathbb{C}$ with Gaussian rational centres and rational radii, we denote by $\tilde{f}$ the restriction of f to the polydisc $D_1 \times \cdots \times D_n$ (and define $\tilde{f}(z)$ to be 0 for $z \notin D_1 \times \cdots \times D_n$). (For $n = 0$, $\tilde{f}$ is taken to be the element f of $\mathcal{H}_0$ ($\subseteq \mathbb{C}$).)

Now, with the usual convention concerning the identification of $\mathbb{C}$ with $\mathbb{R}^2$, we define the structure $\widetilde{\mathbb{R}}(\mathcal{H})$ to be the expansion of $\overline{\mathbb{R}}$ by all such $\tilde{f}$.

Then $\widetilde{\mathbb{R}}(\mathcal{H})$, being a reduct of $\mathbb{R}_{\mathrm{an}}$, is a polynomially bounded, o-minimal expansion of $\overline{\mathbb{R}}$ and its language is countable (since each $\mathcal{H}_n$ is and there are only countably many polydiscs to which we restrict the functions therein).

In [Wilkie 2008] I gave a characterization of the definable, holomorphic functions of $\widetilde{\mathbb{R}}(\mathcal{H})$ around *generic* points of $\mathbb{C}^n$. This characterization has been shown to be insufficient around nongeneric points (see [Jones et al. 2019]), but at least it does give an alternative description of the LD-pregeometry in terms that avoid notions of general definability. The characterization is as follows.

Consider an LD-generic point $a = \langle a_1, \ldots, a_n \rangle \in \mathbb{C}^n$ and let F be a definable, holomorphic function (definable, that is, in the structure $\widetilde{\mathbb{R}}(\mathcal{H})$ without parameters). Then, as is proved in [Wilkie 2008], there exist disks $D_1, \ldots, D_n$ in $\mathbb{C}$ (with centres and radii as specified above) and, for some $N \geq 1$, functions $f_1, \ldots, f_N \in \mathcal{H}_{n+N}$, and definable, holomorphic functions $\phi_1, \ldots, \phi_N : D_1 \times \cdots \times D_n \to \mathbb{C}$ such that

(3) $a \in D_1 \times \cdots \times D_n \subseteq \mathrm{dom}(F)$;

(4) for all $z \in D_1 \times \cdots \times D_n$ we have that $\langle \phi_1(z), \ldots, \phi_N(z) \rangle$ is a nonsingular zero of the map $f_z : \mathbb{C}^N \to \mathbb{C}^N : w \mapsto \langle f_1(z, w), \ldots, f_N(z, w) \rangle$;

(5) $F(z) = \phi_1(z)$ for all $z \in D_1 \times \cdots \times D_n$.

In other words, F arises, at least close to the generic point a, as a coordinate function of a map given by an application of the implicit function theorem applied to functions from $\mathcal{H}$. (Actually, the characterization from [Wilkie 2008, 1.5 and 1.6] makes use of just the one (dependent) variable version of the implicit function theorem together with composition, but it is easy to see that this formulation is equivalent. Note also that the operations 1.2 and 1.3 from [Wilkie 2008] follow from the corresponding closure conditions that we have placed on $\mathcal{H}$. One can consult [Sfouli 2012] for more on this.)

For later use we remark now that (4) is equivalent to

(4*) for all $z \in D_1 \times \cdots \times D_n$ and $j = 1, \ldots, N$ we have $f_j(z, \phi_1(z), \ldots, \phi_N(z)) = 0$ and $J(z, \phi_1(z), \ldots, \phi_N(z)) \neq 0$, where J is the determinant of the Jacobian matrix $(\partial f_j / \partial w_i)_{1 \le i, j \le N}$.

(Note that $J \in \mathcal{H}_{n+N}$.)

In [Wilkie 2008] I define a pregeometry $\widetilde{D}$ on $\mathbb{C}$ associated with implicit functions as discussed above. Namely, a d-tuple $\langle b_1, \ldots, b_d \rangle \in \mathbb{C}^d$ is declared to be $\widetilde{D}$-generic if there do not exist $k \ge 0$ and $b_{d+1}, \ldots, b_{d+k} \in \mathbb{C}$ and $g_1, \ldots, g_{k+1} \in \mathcal{H}_{d+k}$ such that $\langle b_1, \ldots, b_d, b_{d+1}, \ldots, b_{d+k} \rangle$ is a nonsingular zero of the map

$$\langle g_1, \ldots, g_{k+1} \rangle : \mathbb{C}^{d+k} \to \mathbb{C}^{k+1}.$$

It is shown in [Wilkie 2008, Theorem 1.10] that LD and $\widetilde{D}$ are identical pregeometries and as regards to our present aim this leads to our next theorem. It states that one only needs to check that implicitly defined functions, such as the ϕ_i's mentioned above, have analytic continuations along generic paths. To be more precise we make the following:

Definition 5. We say that the suitable sequence $\mathcal{H}$ has the *weak analytic continuation property* (WACP) if for all $n, N \ge 1$, all $a, \omega \in \mathbb{C}^n$ such that $\lambda_{a,\omega}$ is $\widetilde{D}$-generic on $[0, 1]$, all $r \in [0, 1]$ and all $f \in (\mathcal{H}_{n+N})^N$, if $\gamma : [0, r) \to \mathbb{C}^N$ is a continuous map such that $\langle \lambda_{a,\omega}(t), \gamma(t) \rangle \in Z_{\mathrm{reg}}(f)$ for all $t \in [0, r)$ then $\|\gamma(t)\| \nrightarrow \infty$ as $t \to \infty$. (The nonsingularity here is with respect to the last N variables (as in (4*)), and $\|\cdot\|$ is some standard norm on $\mathbb{C}^N$.)

Notice that there is no explicit mention of definability here. Nevertheless, we have the following:

Theorem 6. *If $\mathcal{H}$ has the WACP then $\widetilde{\mathbb{R}}(\mathcal{H})$ has the ACP. Hence, by Theorem 4, the corresponding expansion $\widetilde{\mathbb{C}}(\mathcal{H})$ of the complex field is quasiminimal (even for the language $\mathcal{L}(\widetilde{\mathbb{C}}(\mathcal{H})_{\infty,\omega})$).*

Remark. The structure $\widetilde{\mathbb{C}}(\mathcal{H})$ is defined to be the expansion of the complex field by all l-dimensional, locally definable, complex submanifolds of $\mathbb{C}$ (for all l, n with $1 \le l < n$) with respect to the expansion $\widetilde{\mathbb{R}}(\mathcal{H})$ of the real field. It is trivial to show that all functions in $\mathcal{H}$ are definable in $\widetilde{\mathbb{C}}(\mathcal{H})$.

Proof of Theorem 6. Let $a \in \mathbb{C}^n$ be an LD-generic point and F a definable holomorphic function with $a \in \operatorname{dom}(F)$. Let $\omega \in \mathbb{C}^n$ be such that $\lambda_{a,\omega}$ is generic on $[0, 1]$. We write λ for $\lambda_{a,\omega}$. We must find G satisfying the conclusion of Definition 1.

Choose $\langle D_1, \dots, D_n\rangle$, N, $f = \langle f_1, \dots, f_N\rangle$ and $\phi = \langle \phi_1, \dots, \phi_N\rangle$ as in (3), (4) and (5).

Now we may suppose that *every* zero, $\langle z^{(0)}, w^{(0)}\rangle$ say, of f satisfies

$$J(\langle z^{(0)}, w^{(0)}\rangle) \neq 0$$

(see (4*)). Indeed, let $f_{N+1} \in \mathcal{H}_{n+N+1}$ be defined by

$$f_{N+1}(z, w, w_{N+1}) := w_{N+1} \cdot J(z, w) - 1.$$

Then, letting $f^* := \langle f_1, \dots, f_N, f_{N+1}\rangle$, one easily calculates that the Jacobian of f^* (with respect to $w_1, \dots, w_{N+1}$) has determinant $J(z, w)^2$, which is nonzero whenever $f_{N+1}(z, w, w_{N+1})$ is zero. Further, any nonsingular (with respect to $w_1, \dots, w_{N+1}$) zero of f gives rise to a (unique) zero of the map f^*. So we may replace f by f^* and (3), (4) and (5) remain true (by setting $\phi_{N+1}(z) = J(z, \phi_1(z), \dots, \phi_N(z))^{-1}$). And now, *all* zeros of f^* are nonsingular (with respect to $w_1, \dots, w_{N+1}$).

So we continue our proof with this nonsingularity assumption.

Let T be the set of all those $t \ge 0$ having the following properties:

$(6)_t$ there exists a definable, open, connected set $U_t \subseteq \mathbb{C}$ with $[0, t] \subseteq U_t$;

$(7)_t$ there exists a definable, holomorphic map $\psi^{(t)}$ with range contained in $\mathbb{C}^N$ and with $\lambda(U_t) \subseteq \operatorname{dom}(\psi^{(t)})$ which satisfies the following two conditions;

$(8)_t$ for all $u \in U_t$ we have that $\langle \lambda(u), \psi^{(t)}(\lambda(u))\rangle$ is a nonsingular zero of f (with respect to $w_1, \dots, w_N$);

$(9)_t$ $\psi^{(t)}(\lambda(0)) = f(\lambda(0))$. (So, in particular, the first coordinate of $\psi^{(t)}(\lambda(0))$ is $F(a)$.)

We shall be done if we can show that $1 \in T$, for then we take G to be the first coordinate of $\psi^{(1)}$ to satisfy the conclusion of Definition 1.

Notice that we certainly have $0 \in T$ by taking U_0 to be a (definable) disk around $0 \in \mathbb{C}$ which is small enough to satisfy $\lambda(U_0) \subseteq D_1 \times \cdots \times D_n$, and then taking $\psi^{(0)} := \phi$.

Notice also that if $t_1, t_2 \in T$ and $t_1 \le t_2$ then $\psi^{(t_1)} \circ \lambda$ and $\psi^{(t_2)} \circ \lambda$ must agree on $(U_{t_1} \cap U_{t_2})^*$, the connected component of $U_{t_1} \cap U_{t_2}$ containing the interval $[0, t_1]$.

This is because they agree at 0 (by $(9)_{t_1}$ and $(9)_{t_2}$) and $\lambda(0)$ $(= a)$ is LD-generic, so the definable holomorphic map $(\psi^{(t_2)} - \psi^{(t_1)}) \restriction (U_{t_1} \cap U_{t_2})^*$ must be identically zero (the zero, that is, of $\mathbb{C}^N$).

We now set $r := \sup\{t : [0, t] \subseteq T\}$ and we need to show that $r \geq 1$. So suppose, for a contradiction, that $r < 1$. By the extension property just proved, it follows that $\chi := \bigcup_{t<r}(\psi^{(t)} \restriction \lambda([0, t]))$ is a continuous map with domain $\lambda([0, r))$ such that for all $t \in [0, r)$ we have that $\langle \lambda(t), \chi(\lambda(t)) \rangle$ is a nonsingular zero of f (with respect to $w_1, \ldots, w_N$); see $(8)_t$. So by applying the WACP (with $\gamma = \chi \circ \lambda$) we see that there exists some positive R and an increasing sequence $\langle t_p : p \geq 0 \rangle$ in $[0, r)$ converging to r such that $\|\psi(\lambda(t_p))\| \leq R$ for all $p \geq 0$.

Let $w^{(0)} \in \mathbb{C}^N$ be a limit point of the sequence $\langle \chi(\lambda(t_p)) : p \geq 0 \rangle$. Then $\langle \lambda(r), w^{(0)} \rangle$ is a zero of f (since f is certainly a continuous map throughout $\mathbb{C}^{n+N}$) and by our nonsingularity assumption, it is a nonsingular zero (with respect to $w_1, \ldots, w_N$). So by the implicit function theorem there exist an open polydisc $\Delta \subseteq \mathbb{C}^n$ (which we may take to be definable) with $\lambda(r) \in \Delta$, and a holomorphic map $\theta : \Delta \to \mathbb{C}^N$ satisfying $\theta(\lambda(r)) = w^{(0)}$ and such that for all $z \in \Delta$, the $(n+N)$-tuple $\langle z, \theta(z) \rangle$ is a nonsingular zero of f (with respect to $w_1, \ldots, w_N$). Further, we may assume that Δ has been chosen small enough for there to exist a (definable) open polydisk $E \subseteq \mathbb{C}^N$ with $w^{(0)} \in E$ such that for all $z \in \Delta$, $w = \theta(z)$ is the one and only solution in E of the equation $f(z, w) = 0$. (This follows from the uniqueness condition in the conclusion of the implicit function theorem.) It follows from this that θ is definable.

Now choose p large enough that t_p is close enough to r to satisfy

$$\lambda(t_p) \in \Delta \tag{10}$$

and

$$\chi(\lambda(t_p)) \in E \tag{11}$$

(and hence $\psi^{(t_p)}(\lambda(t_p)) \in E$).

Now choose $\epsilon > 0$ so small that the rectangle

$$\rho := \{x + iy \in \mathbb{C} : t_p - \epsilon < x < t_p + \epsilon, -\epsilon < y < \epsilon\}$$

is contained in U_{t_p}. (See $(6)_{t_p}$.)

Since λ is linear, $\lambda(\rho)$ is a convex, open subset of $\mathbb{C}^n$ $(= \mathbb{R}^{2n})$ and since Δ is too, it follows that $\lambda(\rho) \cap \Delta$ is convex and, in particular, connected.

Now $\lambda(t_p) \in \lambda(\rho) \cap \Delta$ (see (10)) and both $\theta(\lambda(t_p))$ and $\phi^{(t_p)}(\lambda(t_p))$ lie in E (see (11)). So by the uniqueness condition we have that $\phi^{(t_p)}(\lambda(t_p)) = \theta(\lambda(t_p))$ (see $(8)_{t_p}$).

But λ is generic (for either pregeometry) on $[0, 1]$ so, by (2), $\phi^{(t_p)}$ and θ must agree on a sufficiently small, definable open polydisk containing the point $\lambda(t_p)$ and contained within $\lambda(\rho) \cap \Delta$. But then, by the principle of analytic continuation, they must agree throughout the connected set $\lambda(\rho) \cap \Delta$. (Note that $\lambda(\rho) \cap \Delta \subseteq \operatorname{dom}(\phi^{(t_p)}) \cap \operatorname{dom}(\theta)$; see $(7)_{t_p}$.) So we may consistently define a holomorphic map, Γ say, with domain $\lambda(\rho) \cup \Delta$ and taking values in C^N by specifying $\Gamma(z)$ to be $\phi^{(t_p)}(z)$ for $z \in \lambda(\rho)$ and $\theta(z)$ for $z \in \Delta$.

Finally, since $\lambda(r) \in \Delta$ we may choose r' with $r < r' \leq 1$ such that $\lambda(r') \in \Delta$. We now obtain a contradicion by showing that $[0, r'] \subseteq T$.

Indeed, suppose $s \in [0, r']$. Then $\lambda([0, s]) \subseteq \lambda(\rho) \cup \Delta$ (because $\lambda([0, t_p]) \subseteq \lambda(\rho)$ and $\lambda([t_p, s]) \subseteq \Delta$ since Δ is convex). So if we take U_s to be a definable, open, connected subset of $\mathbb{C}$ containing the interval $[0, s]$ and contained within $\lambda^{-1}(\lambda(\rho) \cup \Delta)$, and set $\psi^{(s)} := \Gamma \restriction \lambda(U_s)$, we see that $(6)_s$–$(9)_s$ are all satisfied. Therefore $s \in T$ as required. $\square$

Exponentiation

We now consider the setting appropriate, with respect to the preceding discussion of analytic continuation, for Zilber's quasiminimality conjecture for the complex exponential field.

We let K be a fixed countable, real closed subfield of $\mathbb{R}$ and let $\mathcal{E}_0$ be the algebraically closed subfield $K[i]$ of $\mathbb{C}$. Then for each $n \geq 1$ we let $\mathcal{E}_n$ be the ring $\mathcal{E}_0[z_1, \ldots, z_n, e^{z_1}, \ldots, e^{z_n}]$ of *exponential polynomials* over $\mathcal{E}_0$ in the complex variables $z_1, \ldots, z_n$. It is clear that $\mathcal{E} := \langle \mathcal{E}_n : n \geq 0 \rangle$ is a suitable sequence and one easily checks that $\widetilde{\mathbb{R}}(\mathcal{E})$ is essentially the same (i.e., has the same definable sets) as the structure $\mathbb{R}_K^{\mathrm{RE}}$ considered in [Binyamini and Novikov 2017]. The superscript "RE" stands for "restricted elementary": $\mathbb{R}_K^{\mathrm{RE}}$ is the expansion of $\overline{\mathbb{R}}$ by the restricted functions $\exp \restriction [0, 1]$ and $\sin \restriction [0, \pi]$ and by a constant for each element of K.

Thus, in the present context, a subset M of $\mathbb{C}^n$ is an l-dimensional, locally definable, complex submanifold of $\mathbb{C}^n$ if it satisfies (a) and (b) of Definition 2, where definability is now with respect to the structure $\mathbb{R}_K^{\mathrm{RE}}$ (and, as before, is *without* parameters).

Let us refer to such M simply as *elementary complex manifolds* (ECMs). We write $\mathbb{C}_K^{\mathrm{ECM}}$ for the structure denoted $\widetilde{\mathbb{C}}$ in Definition 3; that is, $\mathbb{C}_K^{\mathrm{ECM}}$ is the expansion of the complex field by all ECMs.

$\mathbb{C}_K^{\mathrm{ECM}}$ is certainly an expansion of $\mathbb{C}_{\exp}$. (See the remark following Definition 3.) Notice also that the notion of an ECM is, at least apparently, more general than if we had required the functions $G_1, \ldots, G_{n-l}$ of Definition 2(b) to lie in $\mathcal{E}_n$. Perhaps the methods of [Jones et al. 2019] could be used to produce an ECM which is not of this latter kind.

Theorem 6 implies:

Theorem 7. *If $\mathcal{E}$ has the WACP then the structure $\mathbb{C}_K^{\mathrm{ECM}}$ (and so, in particular, the structure $\mathbb{C}_{\exp}$) is quasiminimal, even for the language $\mathcal{L}(\mathbb{C}_K^{\mathrm{ECM}})_{\infty,\omega}$.* $\square$

I conclude this article with a proof that $\mathcal{E}$ has the WACP in the case $N = 1$ in the hope that others might find a way to generalize the method, which uses the (algebraic) valuation inequality.

So, suppose that $n \geq 1$ and $a, \omega \in \mathbb{C}^n$ are such that $\lambda := \lambda_{a,\omega}$ is $\widetilde{D}$-generic on $[0, 1]$. Suppose further that $f \in \mathcal{E}_{n+1}$, $r \in [0, 1]$ and that $\gamma : [0, r) \to \mathbb{C}$ is a continuous function such that

$$f(\lambda(t), \gamma(t)) = 0 \neq \frac{\partial f}{\partial z_{n+1}}(\lambda(t), \gamma(t)) \quad \text{for all } t \in [0, r). \tag{12}$$

We must show that $|\gamma(t)| \nrightarrow \infty$ as $t \to r$.

In order to set up a use of the valuation inequality we require the following general fact.

Lemma. *Suppose that $r > 0$ and that $\phi : [0, r) \to \mathbb{C}$ is a continuous function such that $|\phi(t)| \to \infty$ as $t \to r$. Then* ***either***

(A) *for all integers k, l, not both 0 with $k \geq 0$, we have either $|\phi(t)^k \exp(l\phi(t))| \to 0$ as $t \to r$ or $|\phi(t)^k \exp(l\phi(t))| \to \infty$ as $t \to r$,* ***or***

(B) *for some integers k, l, not both 0 with $k \geq 0$, we have that for all countable sets $S \subseteq \mathbb{C}$, there exists $\alpha \in \mathbb{C} \setminus S$ and an increasing sequence $0 \leq t_0 < t_1 < \cdots < t_p < \cdots$ converging to r such that $\lim_{j\to\infty} \phi(t_j)^k \exp(l\phi(t_j)) = \alpha$.*

Proof. Set $J := \{\langle k, l\rangle \in \mathbb{Z}^2 : k \geq 0 \text{ and } k, l \text{ not both } 0\}$. For $\langle k, l\rangle \in J$ write $h_{k,l}(t)$ for $\phi(t)^k \exp(l\phi(t))$ (for $t \in [0, r)$), and define $c^+_{k,l} := \limsup_{t\to r} |h_{k,l}(t)|$ and $c^-_{k,l} := \liminf_{t\to r} |h_{k,l}(t)|$. Then $0 \leq c^-_{k,l} \leq c^+_{k,l} \leq \infty$.

If for each $\langle k, l\rangle \in J$ we have either $0 = c^-_{k,l} = c^+_{k,l}$ or $c^-_{k,l} = c^+_{k,l} = \infty$ then clearly (A) holds. Otherwise, choose $\langle k, l\rangle \in J$ with $c^+_{k,l} > 0$ and $c^-_{k,l} < \infty$. Let S be a countable subset of $\mathbb{C}$. Write h for $h_{k,l}$. Now either $c^-_{k,l} < c^+_{k,l}$ or $0 < c^-_{k,l} = c^+_{k,l} < \infty$.

In the first case choose $c \in \mathbb{R}$ with $c^-_{k,l} < c < c^+_{k,l}$ and $c \notin \{|s| : s \in S\}$. By the continuity of $|h|$ there clearly exists a sequence $0 \leq t'_0 < t'_1 < \cdots$ converging to r such that $|h(t'_j)| = c$ for all $j \in \mathbb{N}$. But now $\langle h(t'_j) : j \in \mathbb{N}\rangle$ is a bounded sequence of complex numbers and hence has a convergent subsequence $\langle h(t_j) : j \in \mathbb{N}\rangle$ whose limit, α say, cannot lie in S because $|\alpha| = c$.

In the second case we have $\lim_{t\to r} |h(t)| = c$, say, with $0 < c < \infty$. There is no harm in assuming that both h and ϕ are nonzero throughout $[0, r)$ and so there exist continuous functions $\theta, \psi : [0, r) \to \mathbb{R}$ such that

$$h(t) = |h(t)| \exp(i\theta(t)), \quad \phi(t) = |\phi(t)| \exp(i\psi(t)),$$

for all $t \in [0, r)$.

So by definition of h we have

$$|h(t)| \exp(i\theta(t)) = |\phi(t)|^k \exp(ik\psi(t)) \cdot \exp\big(l|\phi(t)|(\cos\psi(t) + i \sin\psi(t))\big). \tag{13}$$

Hence $|h(t)| = |\phi(t)|^k \exp\big(l|\phi(t)| \cos(\psi(t))\big) \to c$ as $t \to r$.

We cannot have $l = 0$, for then $k > 0$ and ϕ would be bounded. So $l \neq 0$ from which it follows that $\cos\psi(t) \to 0$ as $t \to r$ (since $c \neq 0, \infty$). Thus ψ is bounded and $\sin\psi(t) \to \pm 1$ as $t \to r$. Equating arguments in (13) we obtain, for some fixed $N_0 \in \mathbb{Z}$ and for all $t \in [0, r)$,

$$\theta(t) = k\psi(t) + l|\phi(t)| \sin\psi(t) + 2\pi N_0.$$

It follows from this that $\theta(t) \to \pm\infty$ (the sign here depending on the eventual sign of $l \sin\psi(t)$) as $t \to r$.

Thus we may choose some $\theta_0 \in \mathbb{R}$ such that $c\exp(i\theta_0) \notin S$ and for which there exists a sequence $0 \leq t_0 < t_1 < \cdots$ converging to r such that $\theta(t_j) = \theta_0 \pmod{2\pi\mathbb{Z}}$ for all $j \in \mathbb{N}$. It follows that $h(t_j) \to c\exp(i\theta_0)$ as $j \to \infty$, and we are done. $\square$

Now returning to the discussion before the statement of the lemma, suppose, for a contradiction, that (12) holds and that $|\gamma(t)| \to \infty$ as $t \to r$. By definition of $\mathcal{E}_{n+1}$ we see that f has the form

$$f(z_1, \ldots, z_n, z_{n+1}) = \sum_{\langle i,j\rangle \in L} P_{i,j}(z_1, \ldots, z_n) z_{n+1}^i \exp(j z_{n+1}) \tag{14}$$

for some nonempty finite set $L \subseteq \mathbb{N}^2$, where $P_{i,j} \in \mathcal{E}_n \setminus \{0\}$ for each $\langle i, j\rangle \in L$. We must have $L \neq \{\langle 0, 0\rangle\}$ by (12).

By the genericity of λ on $[0, r]$ it routinely follows that for all $P \in \mathcal{E}_n \setminus \{0\}$ there exists some $R_P \geq 1$ such that

$$R_P \geq |P(\lambda(t))| \geq R_P^{-1} \quad \text{for all } t \in [0, r]. \tag{15}$$

Let us pass to a nonprincipal ultrapower ${}^\star\mathbb{C}$ of $\mathbb{C}$ (with corresponding ${}^\star\mathbb{R}$, ${}^\star\mathbb{Z}$, ${}^\star\mathbb{N}$). Then the functions $\lambda_1, \ldots, \lambda_n$ (where $\lambda = \langle\lambda_1, \ldots, \lambda_n\rangle$) and all the $P_{i,j}$'s have natural extensions to the ultrapower and (keeping the same notation for the extended functions) (15) remains true for all $t \in {}^\star\mathbb{R}$ with $0 \leq t \leq r$.

For each such t consider the subfield

$$\mathcal{F}_t := \mathcal{E}_0\big(\lambda_1(t), \ldots, \lambda_n(t), \exp(\lambda_1(t)), \ldots, \exp(\lambda_n(t))\big)$$

of ${}^\star\mathbb{C}$. Then by (the extension to the ultrapower of) (15) it follows that $\mathcal{F}_t$ is actually a subfield of the valuation subring $\mathrm{Fin}({}^\star\mathbb{C})$ ($:= \{z \in {}^\star\mathbb{C} : |z| \leq R \text{ for some } R \in \mathbb{R}\}$) of ${}^\star\mathbb{C}$.

By the continuity of each $P \circ \lambda$ (for $P \in \mathcal{E}_n$) it follows that for all $t_1, t_2 \in {}^\star\mathbb{R}$ with $0 \leq t_1, t_2 \leq r$ and satisfying $t_1 \approx t_2$ (i.e., t_1 infinitesimally close to t_2) we have that $P(\lambda(t_1)) \approx P(\lambda(t_2))$ and so the correspondence $t_1 \mapsto t_2$ induces an isomorphism $\mathcal{I}_{t_1,t_2} : \mathcal{F}_{t_1} \to \mathcal{F}_{t_2}$ with $\mathcal{I}_{t_1,t_2}(z) \approx z$ for all $z \in \mathcal{F}_{t_1}$ (and so, in particular, $\mathcal{I}_{t_1,t_2}(z) = z$ for all $z \in \mathcal{E}_0$).

Further, by the continuity of roots of polynomials (see [Harris and Martin 1987] or, perhaps more appropriately in the present context, [Ross 2022]) the map $\mathcal{I}_{t_1,t_2}$ extends to an isomorphism $\widetilde{\mathcal{I}}_{t_1,t_2} : \widetilde{\mathcal{F}}_{t_1} \to \widetilde{\mathcal{F}}_{t_2}$ of the algebraic closures and we still have (for $t_1 \approx t_2$)

$$\widetilde{\mathcal{F}}_{t_1}, \widetilde{\mathcal{F}}_{t_2} \subseteq \mathrm{Fin}({}^*\mathbb{C}) \tag{16a}$$

and

$$\widetilde{\mathcal{I}}_{t_1,t_2}(z) \approx z \quad \text{for all } z \in \widetilde{\mathcal{F}}_{t_1}. \tag{16b}$$

Now choose any $t^* \in {}^*\mathbb{R}$ with $0 < t^* < r$ and $t^* \approx r$. Extend the function $\gamma : [0, r) \to \mathbb{C}$ to the ultrapower and set

$$\mathcal{H}_{t^*} := \widetilde{F}_{t^*}\big(\gamma(t^*), \exp(\gamma(t^*))\big). \tag{17}$$

Then by (12) and (14), $\mathcal{H}_{t^*}$ is a subfield of ${}^*\mathbb{C}$ of transcendence degree at most 1 over $\widetilde{\mathcal{F}}_{t^*}$ and, in fact, exactly 1 because $\gamma(t^*) \notin \mathrm{Fin}({}^*\mathbb{C})$ (and $\widetilde{\mathcal{F}}_{t^*} \subseteq \mathrm{Fin}({}^*\mathbb{C})$). It also follows from this — and the valuation inequality for the valuation on ${}^*\mathbb{C}$ (restricted to $\mathcal{H}_{t^*}$) associated to the valuation subring $\mathrm{Fin}({}^*\mathbb{C})$ of ${}^*\mathbb{C}$ — that $\gamma(t^*)$ and $\exp(\gamma(t^*))$ have $\mathbb{Q}$-dependent valuations which, by an easy saturation argument, implies that (A) of the lemma (with $\phi = \gamma$) cannot hold (back in the standard situation). So (B) holds and we choose integers k, l not both 0 with $k \geq 0$ and take S to be the subset $\widetilde{F}_r$ of $\mathbb{C}$. Let $\alpha \in \mathbb{C} \setminus S$ and $0 \leq t_0 < t_1 < \cdots$ be a sequence converging to r such that $\lim_{j\to\infty} \phi(t_j)^k \exp(l\phi(t_j)) = \alpha$.

Let $p \in {}^*\mathbb{N}\setminus\mathbb{N}$ and $t^* := t_p$. Then $t^* \approx r$, $t^* < r$ and $\alpha^* := \gamma(t^*)^k \exp(l\gamma(t^*)) \approx \alpha$. Now for no $\beta \in \widetilde{\mathcal{F}}_{t^*}$ do we have $\alpha^* \approx \beta$; for this would imply

$$\widetilde{\mathcal{I}}_{t^*,r}(\beta) \approx \beta \approx \alpha^* \approx \alpha \in \mathbb{C} \setminus \widetilde{\mathcal{F}}_r,$$

which is absurd since $\widetilde{\mathcal{I}}_{t^*,r}(\widetilde{\mathcal{F}}_{t^*}) = \widetilde{\mathcal{F}}_r$.

Since $\widetilde{\mathcal{F}}_{t^*}$ is an algebraically closed subfield of ${}^*\mathbb{C}$ contained in $\mathrm{Fin}({}^*\mathbb{C})$ this now implies that $\widetilde{\mathcal{F}}_{t^*}(\alpha^*)$ is also a subfield of ${}^*\mathbb{C}$ contained in $\mathrm{Fin}({}^*\mathbb{C})$ and is of transcendence degree 1 over $\widetilde{\mathcal{F}}_{t^*}$. However, certainly $\widetilde{\mathcal{F}}_{t^*}(\alpha^*) \subseteq \mathcal{H}_{t^*}$. Further, $\gamma(t^*) \in \mathcal{H}_{t^*}$ and $\gamma(t^*)$, being infinite, is transcendental over $\widetilde{\mathcal{F}}_{t^*}(\alpha^*)$, which forces $\mathcal{H}_{t^*}$ to have transcendence degree at least 2 over $\widetilde{\mathcal{F}}_{t^*}$. This contradiction shows that it cannot be the case that $|\gamma(t)| \to \infty$ as $t \to r$ and completes the proof of the WACP in this rather simple situation.

With a little more care one can show that the function γ above has an extension to a definable (in the structure $\mathbb{R}_K^{\mathrm{RE}}$), holomorphic function with $[0, 1] \subseteq \mathrm{dom}(\gamma)$, which (therefore) satisfies (12) for all $t \in \mathrm{dom}(\gamma)$. It is then an easy matter to set up an inductive argument (using the full version of the valuation inequality for polynomially bounded o-minimal structures; see for example [Speissegger 2002]) to establish the WACP for diagonal systems of exponential polynomial equations:

Proposition 8. *$\mathcal{E}$ has the* ***diagonal*** *WACP, i.e., where the map $f = \langle f_1, \ldots, f_N\rangle$ in Definition 5 satisfies the extra condition that $f_j \in \mathcal{E}_{n+j}$ for $j = 1, \ldots, N$. Further, the map γ has an extension to a definable (in $\mathbb{R}_K^{\mathrm{RE}}$), holomorphic function with $[0, 1] \subseteq \mathrm{dom}(\gamma)$.* □

But, unfortunately, this is a *very* special case.

Further remarks on quasiminimality

Boxall [2020] shows that every formula (parameters allowed) of the language $\mathcal{L}(\mathbb{C}_{\exp})$ having the form $\exists \bar{z}(P(\bar{w}, \bar{z}) = 0)$, where $P(\bar{w}, \bar{z})$ is a term of this language (the $\bar{w}, \bar{z}$ being sequences of variables, not necessarily of the same length), is equivalent (in $\mathbb{C}_{\exp}$) to a countable boolean combination of formulas of the form $(\exists \bar{z} \in \mathbb{Q}^m)\phi(\bar{w}, \bar{z})$, where ϕ is a quantifier-free formula of $\mathcal{L}(\mathbb{C}_{\exp})$ (containing no parameters other than those used in P). This immediately implies that sets of the form

$$\pi_1(Z(P)) := \{w \in \mathbb{C} : \exists \bar{z} \in \mathbb{C}^m \, P(w, \bar{z}) = 0\}$$

are either countable or cocountable.

It is worth mentioning here that, even for the case $m = 1$, this is not a property of entire functions in general. For instance, Alexander [1975] complements earlier work of Tsuji [1944] by giving a complete characterization of sets of the form $\pi_1(Z(F))$ for $F : \mathbb{C}^2 \mapsto \mathbb{C}$ an entire function, and this characterization implies that there do exist such F with both $\pi_1(Z(F))$ and $\mathbb{C} \setminus \pi_1(Z(F))$ uncountable. In particular, there is an expansion $\langle \overline{\mathbb{C}}, F \rangle$ of the complex field $\overline{\mathbb{C}}$ by an entire function F of two variables which is not quasiminimal. However, as pointed out by P. Koiran, it is still not known whether there exists a nonquasiminimal such expansion by an entire function of one variable, or even by finitely many entire functions of one variable. On the other hand, at least we do know (using a combination of ideas from [Koiran 2003], [Wilkie 2005] and [Zilber 2005]) that there exists a *transcendental* entire function $f : \mathbb{C} \mapsto \mathbb{C}$ such that the expansion $\langle \overline{\mathbb{C}}, f \rangle$ of $\overline{\mathbb{C}}$ *is* quasiminimal.

References

[Alexander 1975] H. Alexander, "On a problem of Julia", *Duke Math. J.* **42** (1975), 327–332. MR Zbl

[Bays et al. 2010] M. Bays, J. Kirby, and A. J. Wilkie, "A Schanuel property for exponentially transcendental powers", *Bull. Lond. Math. Soc.* **42**:5 (2010), 917–922. MR Zbl

[Bianconi 2005] R. Bianconi, "Undefinability results in o-minimal expansions of the real numbers", *Ann. Pure Appl. Logic* **134**:1 (2005), 43–51. MR Zbl

[Binyamini and Novikov 2017] G. Binyamini and D. Novikov, "Wilkie's conjecture for restricted elementary functions", *Ann. of Math.* (2) **186**:1 (2017), 237–275. MR Zbl

[Boxall 2020] G. Boxall, "A special case of quasiminimality for the complex exponential field", *Q. J. Math.* **71**:3 (2020), 1065–1068. MR Zbl

[Denef and van den Dries 1988] J. Denef and L. van den Dries, "p-adic and real subanalytic sets", *Ann. of Math.* (2) **128**:1 (1988), 79–138. MR Zbl

[Harris and Martin 1987] G. Harris and C. Martin, "The roots of a polynomial vary continuously as a function of the coefficients", *Proc. Amer. Math. Soc.* **100**:2 (1987), 390–392. MR Zbl

[Jones et al. 2019] G. Jones, J. Kirby, O. Le Gal, and T. Servi, "On local definability of holomorphic functions", *Q. J. Math.* **70**:4 (2019), 1305–1326. MR Zbl

[Karp 1964] C. R. Karp, *Languages with expressions of infinite length*, North-Holland Publishing Co., Amsterdam, 1964. MR Zbl

[Koiran 2003] P. Koiran, "The theory of Liouville functions", *J. Symbolic Logic* **68**:2 (2003), 353–365. MR Zbl

[Macintyre and Wilkie 1996] A. Macintyre and A. J. Wilkie, "On the decidability of the real exponential field", pp. 441–467 in *Kreiseliana*, edited by P. Odifreddi, A K Peters, Wellesley, MA, 1996. MR Zbl

[Peterzil and Starchenko 2001] Y. Peterzil and S. Starchenko, "Expansions of algebraically closed fields in o-minimal structures", *Selecta Math.* (*N.S.*) **7**:3 (2001), 409–445. MR Zbl

[Peterzil and Starchenko 2003] Y. Peterzil and S. Starchenko, "Expansions of algebraically closed fields, II: Functions of several variables", *J. Math. Log.* **3**:1 (2003), 1–35. MR Zbl

[Pila and Wilkie 2006] J. Pila and A. J. Wilkie, "The rational points of a definable set", *Duke Math. J.* **133**:3 (2006), 591–616. MR Zbl

[Ross 2022] D. A. Ross, "Yet another proof that the roots of a polynomial depend continuously on the coefficients", preprint, 2022. arXiv 2207.00123v2

[Sfouli 2012] H. Sfouli, "On the elementary theory of restricted real and imaginary parts of holomorphic functions", *Notre Dame J. Form. Log.* **53**:1 (2012), 67–77. MR Zbl

[Speissegger 2002] P. Speissegger, "Lectures on o-minimality", pp. 47–65 in *Lectures on algebraic model theory*, edited by B. Hart and M. Valeriote, Fields Institute Monographs **15**, Amer. Math. Soc., Providence, RI, 2002. MR Zbl

[Tsuji 1944] M. Tsuji, "Theory of meromorphic functions in a neighbourhood of a closed set of capacity zero", *Jpn. J. Math.* **19** (1944), 139–154. MR Zbl

[Wilkie 2004] A. J. Wilkie, "Diophantine properties of sets definable in o-minimal structures", *J. Symbolic Logic* **69**:3 (2004), 851–861. MR Zbl

[Wilkie 2005] A. J. Wilkie, "Liouville functions", pp. 383–391 in *Logic Colloquium* 2000 (Paris), edited by R. Cori et al., Lect. Notes Log. **19**, Assoc. Symbol. Logic, Urbana, IL, 2005. MR Zbl

[Wilkie 2008] A. J. Wilkie, "Some local definability theory for holomorphic functions", pp. 197–213 in *Model theory with applications to algebra and analysis*, vol. 1, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **349**, Cambridge Univ. Press, 2008. MR Zbl

[Zilber 2005] B. Zilber, "Analytic and pseudo-analytic structures", pp. 392–408 in *Logic Colloquium* 2000 (Paris), edited by R. Cori et al., Lect. Notes Log. **19**, Assoc. Symbol. Logic, Urbana, IL, 2005. MR Zbl

Received 14 Feb 2023. Revised 15 Feb 2023.

ALEX J. WILKIE:

wilkie@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

Model Theory
Vol. 3, No. 2, 2024
https://doi.org/10.2140/mt.2024.3.721

Logic Tea in Oxford

Martin Bays and Jonathan Kirby

We assemble some recollections of Boris Zilber from students and others in the Oxford logic group from the time he arrived in Oxford in 1999 to when the Mathematical Institute moved building in 2013, centred around the daily Logic Tea in the St Giles' building common room.

The institution of Logic Tea. When Boris Zilber arrived in Oxford, he instituted a daily Logic Tea in the Mathematical Institute. Gareth Jones recalls:

> When I first arrived in Oxford, it wasn't quite clear whether Alex Wilkie or Boris would be my supervisor. When Boris and I met, he said something very much like "You should come to tea every day at 4pm". So I did, throughout my time in Oxford.

Nick Peatfield was Boris Zilber's first PhD student, arriving in Oxford just before Boris in 1999.

> In that first year, Logic Tea was the main focus of many days for myself and the other four PhD students in model theory in years above me, all being supervised by Alex. We talked about maths and life. I remember Boris talking about the importance of developing a community in an area of mathematics, and over the next few years we had various visitors and more PhD students, leading to a vibrant community at Logic Tea each afternoon.

The institute was then housed in its original 1966 building at 24–29 St Giles, a short distance from its current location in the Andrew Wiles building. The common room was somewhat poorly lit, getting only a little natural light from a window at the front, when that part of the room was not partitioned off as a classroom, and one window looking out to the bike shed area to the rear. There were lighted cabinets containing some mathematical curiosities, and the lighting from these was a necessary supplement to the ceiling lighting. As a result the room was cosy, and the warm glow of the conversation with people leaning in to scrawl on the small whiteboard-topped coffee tables leaves memories akin to a gathering around a winter fireplace.

MSC2020: 01-06.

Many other research groups gathered for tea once or twice a week, before or after their seminars, but the daily frequency of the logic group tea and the number of its adherents were unusual. Indeed, it was a regular occurrence for the circle of chairs to gradually get pushed back as more and more people came to join the logic tea, often to listen to whatever Boris was talking about that day. Assaf Hasson had been a student of Hrushovski before coming to Oxford as a postdoc to work with Boris. He explains:

> Boris is willing to discuss — with the same level of enthusiasm — his ideas with anyone interested: from young grad students, through great mathematicians, physicists (of course), philosophers, or anyone who happened to sit next to him at dinner or during tea time.

As Juan Diego Caycedo recalls,

> Logic Tea was great, you never knew what you'd get: mathematical ideas, stories from the Soviet Union, or something truly random like a breakfast TV show's discussion of how much sunscreen children should use!

Of course, it was not always Boris taking the lead. Gareth Jones recalls that

> Alex and I often talked about cricket or football, depending on the time of year. I've tried to carry on that tradition with my students (with varying levels of success).

Martin Bays explains how central Logic Tea was to the group in his time:

> Logic Tea was the daily ritual of proffering 30p for a cup of barely drinkable coffee, thanking Helen Cullen who always served it, and sitting at a table to hear the latest iteration of Boris's stories and ideas, a ritual which formed the backbone of my time in Oxford. At that time, Zilber's group all worked on a circle of ideas centred on his programme on exponentiation, and Wilkie's group worked on related topics from a different perspective, so bringing us together for tea gave us a crucial overview of what we were collectively working on, how various parts of it were progressing, and where we might go next. I didn't realise at the time that this was anything special, but this unity of purpose is something I've yet to see in any other group. Conversations slipped, largely at Boris's whim, between stories of the past, speculations on politics and philosophy, discussion of random mathematics or physics outside of our expertise, and the more concrete topics of our research, and I suspect that this free flow was a crucial part of a magic which seems impossible to replicate without a Boris.

The oral tradition of mathematics. For Boris, mathematics seems to be primarily an oral tradition. While written mathematics is of course important, it is the discovery and transmission of mathematical ideas in oral form, perhaps supplemented by a diagram and some symbols on a board, which excites him the most, and his excitement was certainly something he communicated to his students. There was only so much space to write on the coffee tables, but the distinction between a tea-time talk and a seminar was not always clear. Tom Foster and others recall

> . . . those seminars about Zariski geometries where at the start Boris would draw a triangle and by the end of the hour that would still be the only thing on the board.

Cecily Crampin was a student of Alex Wilkie who later changed career to law.

> The thing I most remember about Boris, and which has influenced me in one aspect of being a lawyer, is a talk he gave in the 2000 Logic Colloquium in Paris. To my mind, a lot of the talks tried to take the listener through a proof or proofs as if the listener were reading a paper, and without the time I needed to go back and scratch my head and reread the first bit so I understood the steps. Boris didn't do that. He gave a talk in which he explained why he was interested in the area of logic which he was working on (and in which my DPhil work was based by his suggestion). The description was much more ideas dancing than the nitty gritty of proof. I could see it and I was excited by the vision of what could be achieved. I still remember that, though the maths has long escaped me. That way of speaking about a very technical subject when giving a talk has influenced the way I approach giving seminars in property law.

One possible reason for this emphasis on oral mathematics comes from his experience of undergraduate exams in Russia, which were usually oral exams. One anecdote Boris would recount of those days was of a professor who found a clever way to reduce the time taken to do individual oral exams for the dozens of students in his class. First, he challenged the students that if they had not done the work to prepare, they would be wasting their time to wait for their turn, so he invited those students to leave. Once they had gone, he asked which of the remaining students would be satisfied with a D, the lowest passing mark. They were invited to give their names and leave. Then he asked who was sure they would get an A. No point wasting time on them. Finally, of the remaining students, he asked which of them wanted a chance to upgrade their C to a B, and these he would examine!

This oral mathematical background is visible in the draft manuscripts of Boris with which his students would grapple. He would start with an outline of the argument, and then in places the details of statement and argument and definition would be sketched in more and more detail, as if in answer to an interlocutor.

Tom Foster was one of Alex Wilkie's students who remembers one discussion at Logic Tea

> ...about "proof", where I made what I thought was the uncontroversial point that I didn't think you could know something to be true until you had written down the proof. Of course Boris completely rejected this point. I guess Boris just discovered things in his head and it was up to others to formalise the proof.

Alex himself provided a wonderful counterbalance to some of Boris's traits, and it was perhaps just as well that at Logic Tea we also heard his point of view on proof and could take our own path between the two. Many readers will have seen one of Alex's handwritten manuscripts where arguments are developed carefully, line-by-line.

Another student of Alex Wilkie was Henry Braun, who was already working on Boris's conjecture that the complex exponential field should be quasiminimal when Boris arrived in Oxford. Most memorable to him was

> ...not tea but a chat as we walked together across Oxford to lunch at Merton. Boris observed that the thinking he did indoors and outdoors was different. Indoors, looking close to his hands, was good for manipulative algebra. Outdoors, with eyes raised to the scenery, was the place for geometry. One must be alert to both. The importance of choosing the right environment for the work in hand is something I have carried into my career outside logic.

Russian / British culture. The culture shock of arriving in Oxford from post-Soviet Kemerovo must have been significant, and in the early years this was a regular topic of conversation.

Henry Braun recalls that, in his first couple of weeks, Boris had to get used to being "Boris" to the graduate students, not "Boris Iosefevich", without disrespect.

Nick Peatfield remembers

> ...his recollections from Russia, and in particular his assertion that "if everybody around you keeps repeating something for years and years, you will end up believing it". A scary thought in these times, but also a sign that Boris himself was adjusting to some of the assumptions about life prevalent in the West, but very different to some of those held in the former Soviet Union.

Today, the logic group in Oxford is truly international. However, in the early 2000s almost everyone in the logic group was British. Apart from Boris, just Alex Wilkie's student Mario Edmundo and Boris's student Misha Gavrilovich were the exceptions at that time. As such, Boris's descriptions of a Russia which was, to us, a very alien world were fascinating.

Vinesh Solanki describes how Boris

> ... somehow had the ability to speak matter-of-factly and even lightly about some quite difficult things. I remember when he told us the story about his assignment to a university after he had finished his doctorate (which apparently happened automatically and at random in Soviet Russia). His friend had been assigned to a university that was close to the border with China and the two of them were going to be separated by a considerable distance. I can't quite remember what Boris did to change this state of affairs but it was the way he put in his inimitable and matter-of-fact style. "Anyway, at some point I was told that some random redistribution had occurred which put us close enough to each other; what are the odds?"

He also told awful stories about his father's experience during the Siege of Leningrad and of the constant fear of life under Stalin, which really brought this part of history to life for those of us fortunate enough to have led more sheltered lives.

Vinesh continues:

> Boris was often excited by new mathematics and ideas and his excitement rubbed off on me. As a student, at a university with plentiful resources and good mathematicians, I wished to learn lots of new mathematics and there were no obstacles to doing so. During a conversation with Boris, I remember him telling me that original research was harder to come by in Russia and that he had to be more reliant on his own resources. It was these kinds of stories that allowed me to come to a better understanding of what he valued and that in a number of ways, Boris came from a world that was harder than I understood.

If Soviet Russia was strange to us students, then Oxford was also strange to Boris. He was rather surprised but impressed by the British sense of fairness. He recalled that he had asked Roger Heath-Brown, then Director of Graduate Studies, if the money left over from the graduate student allocation could be spent on this or that student? The answer was no, because it would not be a fair process. It was apparently better for the money to be unspent and no one to benefit than for one person to benefit unfairly!

Boris Zilber's approach to mathematics. Boris has always brought outside ideas into model theory. If algebraic geometry was the source in his early days in Russia, then physics was often the source in Oxford. Much remembered by many students was the occasion when Boris was teaching a course on Zariski geometries when a

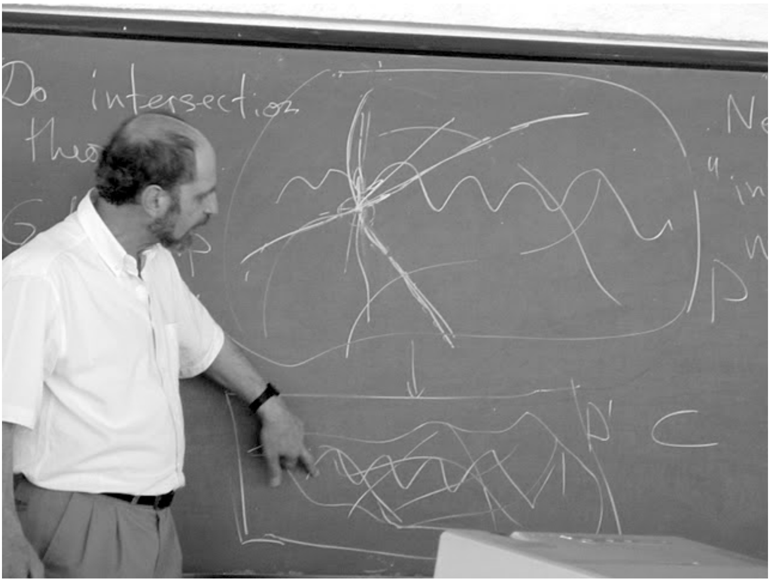

Figure 1. Geometric thinking on the board: a common sight both at Logic Tea and in seminars. Photograph by Andrés Villaveces.

student opened the door from outside and stayed there looking around to decide if this was his lecture. Boris then noticed him, stopped the lecture and asked him what he was looking for. The student asked “Is this quantum?” After a little thought, Boris answered with a smile on his face: “Not yet!”

Jonathan Kirby recalls that at one Logic Tea, Boris asked him to borrow a popular physics book from a college library. Boris soon graduated to the research literature, and his model-theoretic perspective on explaining the mysteries of quantum mechanics via noncommutative geometry became a regular topic at Tea. Indeed, so regular that Assaf Hasson explains:

> The number of times I have seen Boris draw his image of the Zariski geometry he used to interpret as the quantum group at roots of unity is barely countable.

Nick Peatfield recalls:

> One idea of Boris’s that stands out is his description of the time taken to engage in mathematics, in comparison with that taken by empirical scientists in the lab. He said that, whereas those scientists go to work each day and get on with their lab work, we mathematicians have to be “at work” a lot more of the time. We have to be “with” the problem we are working on almost 24 hours a day, as the breakthrough is just as likely to happen when you are brushing your teeth in the morning as when you are

> sitting at your desk consciously mulling it over. This idea has seeped into my work as a mathematics teacher educator, where a phrase I repeat often is that expressed by the mathematics educator John Mason: “Teaching takes place in time; but learning takes place over time”.

Adam Harris and others remember:

> …one day when we were in Pisa at a conference and I asked him what he’d been doing today. He told me that he’d been looking round the sights in Pisa with his wife, thinking about mathematics, pretending not to think about mathematics…

Boris’s mathematical insights were often startling but not easy to grasp. Ayhan Günaydin was a postdoc in Oxford.

> I was working on the expansions of the ordered real field by a finitely generated multiplicative subgroup G of the circle S. While proving an elimination of imaginaries result with Alex Berenstein and Clifton Ealy, we had the idea that the structure induced on the quotient S/G from that expansion could be stable. We had some good reasons to think so! Possibly my first day of working with Boris, he asked about what I had been doing in those days. After presenting some results I had proven, I also mentioned that belief of ours. Boris’s immediate reaction was “but there are curves there!” while twisting his finger in the air. I had that single sentence in my mind for the rest of the day: there are curves there. What could it have meant? Even if there were curves, why would that be opposed to being stable? There are curves in algebraically closed fields anyway…. I had given up trying to understand it and asked him what he meant. He sent me a very old-looking write-up where it is proven that there are certain curves in certain quotients, which gives rise to defining the real field. After that incident, I learned to pay attention to Boris’s fantastical sounding remarks. They are always filled with a lot of insight, however most of the time, it is very hard for us to decipher what he means. It generally takes a great mind like his to do that!

Adam Harris:

> I was often in awe of the way he thought about maths and geometry. There were many times when the conversation moved towards a seemingly far away subject area and he could so easily relate it back to his domain (logic / model theory / stability) in a seemingly simple and obvious way.
>
> I also remember a few times in supervisions when he explained a concept that I believed I understood fairly well in general, but he explained it in such a way that was so illuminating but simple that it immediately

> changed the way I thought about it. There were obviously quite a few fairly confusing times too when he drew another blob with a wiggly line in it and I had no idea what he was talking about!

Boris always made connections and could see past the details to the big picture. Assaf Hasson remembers Boris producing a wonderful phrase:

> I once asked Boris about a preprint I knew he was reading. He said "The results are really very nice, but the paper is locally everywhere wrong". That is still a description I like of a certain kind of badly written papers.

Assaf continued:

> Boris is one of a very few mathematicians I know who has a spark in the eye when talking — not about his results — rather about his mathematical philosophy, vision and conjectures. I think that the Zilber trichotomy is indicative of his style: it is bold, far-reaching, morally true. The fact that it is, in a precise mathematical sense, very wrong does not deter him.

Piotr Kowalski was a postdoc whose time in Oxford overlapped with Assaf Hasson. They were working on Peterzil's conjecture that the Zilber trichotomy should hold for strongly minimal sets which are interpretable in o-minimal structures, and did eventually prove a special case with a long and complicated argument. Piotr recalls:

> Assaf and I were then usually working whole days on this. It turned out to be much more difficult than we had thought and often at the Logic Tea we were discussing the ideas which had collapsed around us. In such moments, Boris was kindly suggesting "Maybe it is time to work on a counterexample?"

The later years. Sylvy Anscombe and Franziska Jahnke, two students of Jochen Koenigsmann, give their own reflections on their experience of Logic Tea with Boris Zilber in the years just before the Mathematical Institute moved buildings.

> Before either of us had ever met Boris, we heard many a legend about him, mostly around his exceptional and deep insights and his central role in shaping our community's research directions through his theorems and conjectures. Long before our time, his reputation preceded him across the iron curtain, and growing up as mathematicians, we heard him spoken of by our professors in awed tones. Martin Ziegler spoke of Boris running an arboretum in Oxford, planting seedlings of ideas into students and colleagues alike. Dugald Macpherson told of the excitement in the Western community when Boris "came to the West". The Oxford administration seemed not as impressed as the research community and

required the award of a fresh Master of Arts degree before he could start his tenure as the chair of mathematical logic.

Personally, we were each interviewed by him before we were accepted as PhD candidates at Oxford, though neither of us was asked testing questions during said interviews. His maxim — then as now — is that it is easy to find a teacher who will explain essentially all they know, but hard to find a student willing to listen. He is a strong believer in learning by permeation, and strongly advocates listening with full attention to talks on almost any subject, in the expectation that frequent exposure to an abstract concept will foster understanding. He led us by example, always willing to ask questions, from those "stupid ones" we were thinking but not daring to ask, to such insightful ones that they turned the speaker's own perspective. At the same time, he freely admitted to a certain nervousness when outlining a research question in front of Saharon Shelah — in case the answer was already obvious to the latter. This humility and his kindness (*Please*! *Call me Boris. My name is Boris*! *Not Prof. Zilber.*) made him approachable from day one. Despite not being supervised by him, we were hugely influenced by his leadership of the Oxford Logic Group.

Although on paper the main event of the week was Logic Seminar, initiates knew that it was Logic Tea that mattered most (incidentally, a daily occurrence with tea served by Helen). Attendance was effectively mandatory, and Boris would gladly pay for your tea if you were short on change. To his lament, the stamina of graduate students for the consumption of tea decreased generation after generation, and eventually only Thursdays would draw a full house. Nevertheless, reliably and daily at 4pm, Boris would be found in the common room of the old Mathematical Institute. It is easy to picture the scene even now: Boris, sharply dressed, trim, sometimes entering the room with his trademark beret.

The topics at tea were as varied as his mathematical interests, and very certainly not limited to mathematics. We learned about life in the Soviet Union, beginning with the stories of his parents during and after the war (as a recommendation: when in the Red Army, do not complain about the lack of shoes), the necessity of bribing officials for just about anything, as well as having Yiddish as a first language (with a smattering of German thrown in for good measure). His life at university brought dangers and adventures: a mandatory undergraduate hike resulted in him and a group of fellow students getting lost in a vast forest, only finding their way back to civilisation by a stroke of luck and a fortuitously located cabin. Later, he had to overcome immense difficulties to obtain a doctorate as a Jewish

> student. Although these were difficult topics, Boris recounted them with lightness and his characteristic good humour.
>
> Meanwhile, the denizens of Logic Tea were as varied as the conversation, and Boris's willingness to give advice extended from the youngest master's student to the most eminent visiting speaker. His generosity in sharing his ideas, both on mathematical and more mundane matters, meant that the conversation was free flowing. Further recurrent themes included the beauty and uses of Zariski geometries, the characteristic of the universe (presumably finite), the mysteries of quantum physics, the reticence of (pseudo)exponentiation, and whether the food or the intricacy of the tables was to be preferred at Bangkok House. An all time favourite of ours is the story of how he and his son willingly let a con-artist at a market change a large note into smaller ones for them, knowing they were about to be duped but still keen to watch how the trick was done.
>
> Let's be categorical: Logic Tea was a logically perfect structure. And if we ever need someone trustworthy with cunning and plenty of life experience to steal a horse with, we will choose Boris.

Boris Zilber's character. While Boris's approach to mathematics is most evident at Logic Tea, many students reflected on his character more widely.

Nick Peatfield recalls:

> Boris welcomed me into the model-theoretic community with care and compassion, and his warm and welcoming personality made me feel at home in a place that was new to both of us.

From Tom Foster:

> I have lots of great memories of my time in the logic group and Boris in particular. He created a very friendly and collaborative atmosphere which at the time I think I just took for granted but which in hindsight I realise was quite special.

Vinesh Solanki reflected similarly:

> If I reflect on Boris, I recall his charisma, wit, strength of intuition and generosity. In my time since leaving Oxford, I have come to realise that this combination of traits is fairly rare.

Adam Harris:

> Overall he is a kind-hearted, brilliant human being and I feel very privileged to have spent so much time around him.

This reputation evidently went beyond Oxford. Juan Diego Caycedo writes:

> Shortly before going to Oxford, I met a few model-theorists at a conference in Colombia. When I mentioned that I was going to be Boris's student, the first thing that they said about him was always the same: that he's such a kind person (not something about mathematical vision or the like). After my years there, that's also the first thing that I say myself when I get a chance.

And, lest we get too caught up in the importance of our research, Juan Diego Caycedo recalls Boris giving some perspective on life:

> Of all the teatime stories, the one that I remember the most is the one about how Boris once made a replacement heel for his wife's shoe. She was upset that her shoe had broken and there seemed to be no way to get it repaired or get a new pair. Boris told us that he put a lot of work into it and that in the end he had been as proud of that heel as of any of his theorems.

Received 6 Jun 2024.

MARTIN BAYS:
bays@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

JONATHAN KIRBY:
jonathan.kirby@uea.ac.uk
School of Mathematics, University of East Anglia, Norwich, United Kingdom

Guidelines for Authors

Authors may submit manuscripts in PDF format on-line at the submission page.

Originality. Submission of a manuscript acknowledges that the manuscript is original and and is not, in whole or in part, published or under consideration for publication elsewhere. It is understood also that the manuscript will not be submitted elsewhere while under consideration for publication in this journal.

Language. Articles are usually in English or French, but articles written in other languages are welcome.

Required items. A brief abstract of about 150 words or less must be included. It should be self-contained and not refer to bibliography keys. If the article is not in English, two versions of the abstract must be included, one in the language of the article and one in English. Also required are keywords and a Mathematics Subject Classification for the article, and, for each author, affiliation (if appropriate) and email address.

Format. Authors are encouraged to use LaTeX and the standard amsart class, but submissions in other varieties of TeX, and exceptionally in other formats, are acceptable. Initial uploads should normally be in PDF format; after the refereeing process we will ask you to submit all source material.

References. Bibliographical references should be complete, including article titles and page ranges. All references in the bibliography should be cited in the text. The use of BibTeX is preferred but not required. Tags will be converted to the house format, however, for submission you may use the format of your choice. Links will be provided to all literature with known web locations and authors are encouraged to provide their own links in addition to those supplied in the editorial process.

Figures. Figures must be of publication quality. After acceptance, you will need to submit the original source files in vector graphics format for all diagrams in your manuscript: vector EPS or vector PDF files are the most useful.

Most drawing and graphing packages — Mathematica, Adobe Illustrator, Corel Draw, MATLAB, etc. — allow the user to save files in one of these formats. Make sure that what you are saving is vector graphics and not a bitmap. If you need help, please write to graphics@msp.org with as many details as you can about how your graphics were generated.

Bundle your figure files into a single archive (using zip, tar, rar or other format of your choice) and upload on the link you been provided at acceptance time. Each figure should be captioned and numbered so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript with all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables.

White Space. Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Proofs. Page proofs will be made available to authors (or to the designated corresponding author) at a Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

Model Theory

no. 2 vol. 3 2024

Special issue on the occasion of the 75th birthday of
Boris Zilber

Introduction 199
MARTIN BAYS, MISHA GAVRILOVICH and JONATHAN KIRBY

Meeting Boris Zilber 203
WILFRID HODGES

Very ampleness in strongly minimal sets 213
BENJAMIN CASTLE and ASSAF HASSON

A model theory for meromorphic vector fields 259
RAHIM MOOSA

Revisiting virtual difference ideals 285
ZOÉ CHATZIDAKIS and EHUD HRUSHOVSKI

Boris Zilber and the model-theoretic sublime 305
JULIETTE KENNEDY

Approximate equivalence relations 317
EHUD HRUSHOVSKI

Independence and bases: theme and variations 417
PETER J. CAMERON

On the model theory of open generalized polygons 433
ANNA-MARIA AMMER and KATRIN TENT

New simple theories from hypergraph sequences 449
MARYANTHE MALLIARIS and SAHARON SHELAH

How I got to like graph polynomials 465
JOHANN A. MAKOWSKY

La conjecture d'algébricité, dans une perspective historique, et surtout modèle-théorique 479
BRUNO POIZAT

Around the algebraicity problem in odd type 505
GREGORY CHERLIN

Finite group actions on abelian groups of finite Morley rank 539
ALEXANDRE BOROVIK

Zilber's skew-field lemma 571
ADRIEN DELORO

Zilber–Pink, smooth parametrization, and some old stories 587
YOSEF YOMDIN

The existential closedness and Zilber–Pink conjectures 599
VAHAGN ASLANYAN

Zilber–Pink for raising to the power *i* 625
JONATHAN PILA

Zilber's notion of logically perfect structure: universal covers 647
JOHN T. BALDWIN and ANDRÉS VILLAVECES

Positive characteristic Ax–Schanuel 685
PIOTR KOWALSKI

Analytic continuation and Zilber's quasiminimality conjecture 701
ALEX J. WILKIE

Logic Tea in Oxford 721
MARTIN BAYS and JONATHAN KIRBY