Model Theory

Celebratory issue on the occasion of

Ehud Hrushovski's 60th Birthday

edited by Assaf Hasson, H. Dugald Macpherson and Silvain Rideau-Kikuchi

no. 2 vol. 2 2023

$M \not\models T$

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. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.133

Introduction

This special issue of *Model Theory* is in honour of Ehud Hrushovski (Oxford University). A conference to celebrate Hrushovski's 60th birthday, entitled "Model theory: from geometric stability to tame geometry", was originally scheduled for 8–12 June 2020 at the CIRM, Luminy, France. Owing to the Covid-19 pandemic, it was postponed until 13–17 December 2021, where it was held as a hybrid workshop of the same title at the Fields Institute in Toronto, as part of the Fields Institute *Thematic Program on Trends in Pure and Applied Model Theory*. This special issue arose from that workshop, with all workshop speakers invited to contribute an article.

Over almost 40 years, Hrushovski's influence on model theory, and on its relationships to other parts of mathematics, has been phenomenal. We cannot here do justice to this, but pick out a few highlights where his highly original contributions have astonished our community and spawned riches for many. We do not touch on his most recent work, which has comparable potential but is less widely absorbed.

Throughout his career, starting with his Berkeley Ph.D. thesis, geometric stability theory (definable groups, minimal types, internality and binding groups, orthogonality, canonical bases, imaginaries, often in unstable contexts) has been a guiding theme; an example is his 2000 paper with Hart and Laskowski *The uncountable spectra of countable theories*, a culmination of classification theory over countable languages. "Hrushovski constructions" first appeared in his talks in 1988 to give counterexamples to conjectures of Lachlan and Zilber but the ideas since then have yielded countless other important examples, as well as Zilber's pseudo-exponential field. His 1996 paper *Zariski geometries* with Zilber exhibits a natural context where Zilber's trichotomy conjecture holds (so is a counterpart to the Hrushovski constructions), and was a key ingredient to his subsequent work on diophantine geometry. Early drafts of his monograph with Cherlin, *Finite structures with few types*, as well as work on PAC structures, gave versions of the independence theorem which underpins simple theories.

On the more applied side of model theory, Hrushovski's work with Chatzidakis and Peterzil on ACFA opened up difference algebra as an area of model-theoretic applications, reinforced by his manuscript on the nonstandard Frobenius. Hrushovski startled not just the model theory community when he found applications of large

Keywords: Ehud Hrushovski, introduction to special volume.

chunks of geometric stability theory in diophantine geometry: he obtained for example a model-theoretic proof of the geometric Mordell–Lang conjecture, in all characteristics (a new result for function fields in characteristic p), and a new proof of the Manin–Mumford conjecture, with explicit bounds. A series of joint papers with Pillay in the late 2000s (one paper also with Peterzil, another also with Simon) yielded new ways of thinking about NIP theories model-theoretically, proved the Pillay conjecture on definable groups in o-minimal expansions of ordered fields, exhibited the significance of Keisler measures, and found model-theoretic applications of Vapnik–Chervonenkis theory; it has yielded many further developments in definable dynamics by Hrushovski and others. His 2009 paper on approximate subgroups led to the Breuillard–Green–Tao classification of finite approximate subgroups and to many other riches. Over the last 25 years, the model theory of valued fields has been a major theme of his work, with ideas from stability theory feeding into our understanding of algebraically closed valued fields, with applications in motivic integration (Hrushovski and Kazhdan), nonarchimedean tame topology and Berkovich space (Hrushovski and Loeser), and zeta functions for groups (Hrushovski, Martin, and Rideau-Kikuchi).

We do not here comment on all the articles in this issue, but note how several of them reflect or grew out of Hrushovski's work.

The paper *Residue field domination in some henselian valued fields* by C. Ealy, D. Haskell, and P. Simon builds directly on the monograph *Stable domination and independence in algebraically closed valued fields* by Haskell, Hrushovski, and Macpherson, which developed "stable domination" as an abstraction of the way, in algebraically closed valued fields, certain types (e.g., the generic type of the valuation ring) are governed by their trace in the residue field. The analogue to Berkovich analytification developed by Hrushovski and Loeser was the space of stably dominated types concentrating on a variety. The Ealy–Haskell–Simon paper develops an analogue of stable domination for other henselian valued fields.

Kamensky's *Higher internal covers* grew from Hrushovski's influential paper *Groupoids*, *imaginaries*, *and internal covers*.

Remarks around the nonexistence of difference-closure by Chatzidakis shows that difference fields do not in general have a difference closure, but develops a stronger notion of κ-closure which, under extra hypotheses, exists and is unique up to isomorphism. This paper is a natural development of the body of work initiated by Chatzidakis and Hrushovski (one paper also with Peterzil) on the theory ACFA, the model companion of the theory of difference fields.

Breuillard's *An exposition of Jordan's original proof of this theorem on finite subgroups of* $\mathrm{GL}_n(\mathbb{C})$ is in part historical. It has connections to a number of important themes in Hrushovski's work: finite approximate subgroups, pseudofinite dimension (originating in a paper of Hrushovski and Wagner), and a model-theoretic approach to the Larsen–Pink strengthening of Jordan's theorem, using the "dimension comparison lemma" of Hrushovski–Wagner.

In *Rigid differentially closed fields*, Marker constructs differentially closed fields of characteristic 0 with no nontrivial automorphisms. The construction makes essential use of the Hrushovski–Sokolovic analysis of strongly minimal sets in differentially closed fields, itself a key ingredient for his work on Mordell–Lang.

Higher amalgamation properties in measured structures by Evans builds on several themes in Hrushovski's work. Higher amalgamation is central in Hrushovski's "Groupoids" paper, and this paper by Evans focuses on ω-categorical Hrushovski constructions as a source of examples. Measures in model theory are also a persistent theme in Hrushovski's work, for example in his paper on approximate subgroups, in the above-mentioned NIP papers, and in much more recent work.

Measures in model theory are also central to *Definable convolution and idempotent measures, II* by Chernikov and Gannon. This article develops many ideas from the NIP papers of Hrushovski and Pillay, also involving Peterzil and Simon.

We believe the articles in this special issue are both important in their own right, and a fitting tribute to Hrushovski's impact on model theory and its applications.

The Editors

ASSAF HASSON:
hassonas@math.bgu.ac.il
Ben Gurion University of the Negev, Beer Sheva, Israel

DUGALD MACPHERSON:
h.d.macpherson@leeds.ac.uk
School of Mathematics, University of Leeds, Leeds, United Kingdom

SILVAIN RIDEAU-KIKUCHI:
silvain.rideau-kikuchi@ens.fr
Département des mathématiques et applications, École normale supérieure, Paris, France

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.137

Mock hyperbolic reflection spaces and Frobenius groups of finite Morley rank

Tim Clausen and Katrin Tent

We define the notion of mock hyperbolic reflection spaces and use it to study Frobenius groups. These turn out to be particularly useful in the context of Frobenius groups of finite Morley rank including the so-called *bad groups*. We show that connected Frobenius groups of finite Morley rank and odd type with nilpotent complement split or interpret a bad field of characteristic zero. Furthermore, we show that mock hyperbolic reflection spaces of finite Morley rank satisfy certain rank inequalities, implying, in particular, that any connected Frobenius group of odd type and Morley rank at most ten either splits or is a simple nonsplit sharply 2-transitive group of characteristic $\neq 2$ of Morley rank 8 or 10.

1. Introduction

This paper contributes to the study of groups acting on geometries arising naturally from conjugacy classes of involutions. We define the notion of a mock hyperbolic reflection space and use it to study certain Frobenius groups. Such an approach to the classification of groups and their underlying geometries based on involutions was developed by Bachmann [1959]. Mock hyperbolic reflection spaces generalize real hyperbolic spaces and their definition is motivated by the geometry arising from the involutions in certain nonsplit sharply 2-transitive groups.

MSC2020: 03C45, 03C60, 20B22, 20F11.
Keywords: Frobenius groups, sharply 2-transitive groups, groups of finite Morley rank, Bachmann geometries.

The points of such a mock hyperbolic space are given by a conjugacy class of involutions, and we view the conjugation action by an involution in the space as a point-reflection. More precisely, a conjugacy class of involutions in a group forms a *mock hyperbolic reflection space* if it admits the structure of a linear space such that three axioms are satisfied: three points are collinear if and only if the product of their point-reflections is a point-reflection, for any two points there is a unique midpoint, i.e., a unique point reflecting one point to the other, and given two distinct lines there is at most one point reflecting one line to the other.

We will consider in particular mock hyperbolic reflection spaces arising from Frobenius groups of finite Morley rank. One of the main open problems about groups of finite Morley rank is the *algebraicity conjecture*, which states that any infinite simple group of finite Morley rank should be an algebraic group over an algebraically closed field. While the conjecture was proved by Altınel, Borovik, and Cherlin [Altınel et al. 2008] in the characteristic 2 setting, it is still wide open in general and in particular in the situation of small (Tits) rank. The conjecture would in fact imply that any sharply 2-transitive group of finite Morley rank and, more generally, any Frobenius group of finite Morley rank splits.

A Frobenius group is a group G together with a proper nontrivial malnormal subgroup H, i.e., a subgroup H such that $H \cap H^g = \{1\}$ for all $g \in G \setminus H$. (If G is a *bad group* of finite Morley rank with Borel subgroup B then $B < G$ is a Frobenius group.) A classical result due to Frobenius states that finite Frobenius groups split, i.e., they can be written as a semidirect product of a normal subgroup and the subgroup H. In the setting of finite groups the methods used by Frobenius play an important role in the classification of CA-groups, CN-groups, and groups of odd order. For groups of finite Morley rank, all the corresponding classification problems are still wide open.

Sharply 2-transitive groups of finite Morley rank came to renewed attention when recently the first sharply 2-transitive groups without nontrivial abelian normal subgroup were constructed in characteristic 2 in [Rips et al. 2017] (see also [Tent and Ziegler 2016]) and in characteristic 0 in [Rips and Tent 2019]. However, as we show below, these groups do not have finite Morley rank. We also show that specific nonsplit sharply 2-transitive groups of finite Morley rank would indeed be direct counterexamples to the algebraicity conjecture.

We prove the following splitting criteria for groups with an associated mock hyperbolic reflection space:

Theorem 1.1. *If G is a group with an associated mock hyperbolic reflection space J, then the following are equivalent*:

(a) $G \cong A \rtimes \mathrm{Cen}(q)$ *for some abelian normal subgroup A and any $q \in J$.*

(b) *J is a* (*possibly degenerate*) *projective plane.*

(c) *J consists of a single line.*

We show a rank inequality for mock hyperbolic reflection spaces in groups of finite Morley rank: if J is a mock hyperbolic reflection space of Morley rank n such that lines are infinite and of Morley rank k, then $n \leq 2k$ implies that J consists of a single line (and hence $n = k$). If $n = 2k + 1$, then there exists a normal subgroup similar to the one in the above theorem (see Theorem 6.11).

We then consider mock hyperbolic reflection spaces arising from Frobenius groups. A connected Frobenius group G of finite Morley rank with Frobenius complement H falls into one of three classes: it is either degenerate or of odd or of even type depending on whether or not G and H contain involutions (see Section 4). A connected Frobenius group is of odd type if and only if the Frobenius complement contains an involution. In particular, every sharply 2-transitive group of finite Morley rank and characteristic different from 2 is a Frobenius group of odd type. We show:

Theorem 1.2. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type.*

(a) *The involutions J in G form a mock hyperbolic reflection space and all lines are infinite.*

(b) *If a generic pair of involutions is contained in a line of Morley rank k and* $\mathrm{MR}(J) \leq 2k + 1$, *then $H < G$ splits.*

(c) *If G does not split and a generic pair of involutions is contained in a line of Morley rank* 1, *then G is a simple sharply* 2*-transitive group of characteristic* $\neq 2$ *and hence a direct counterexample to the algebraicity conjecture.*

(d) *If* $\mathrm{MR}(G) \leq 10$, *then either G splits or G is a simple nonsplit sharply* 2 *transitive group of characteristic* $\neq 2$ *and* $\mathrm{MR}(G)$ *is either* 8 *or* 10.

For nilpotent Frobenius complements we show the following splitting criteria:

Theorem 1.3. *If $H < G$ is a connected Frobenius group of finite Morley rank and odd type with nilpotent complement H, then any of the following conditions implies that $H < G$ splits*:

- *H is a minimal group.*
- *The lines in the associated mock hyperbolic reflection space have Morley rank* 1.
- *G does not interpret a bad field of characteristic* 0.

If G is a uniquely 2-divisible Frobenius group, then G does not contain involutions. However if the complement H is abelian, then we can use a construction from the theory of K-loops to extend G to a group containing involutions and if $H < G$ is full, i.e., if $G = \bigcup_{g \in G} H^g$, then the involutions in this extended group will again form a mock hyperbolic reflection space (see Section 4).

This construction allows us to use mock hyperbolic reflection spaces to study Frobenius groups of finite Morley rank and degenerate type. This class contains potential bad groups. Frécon [2018] showed that bad groups of Morley rank 3 do not exist. Subsequently, Wagner [2017] used Frécon's methods to show more generally that if $H < G$ is a simple *full* Frobenius group of Morley rank n with abelian Frobenius complement H of Morley rank k, then $n > 2k + 1$. Note that the existence of full Frobenius groups was claimed by Ivanov and Olshanski, but to the authors' best knowledge no published proof exists (see also [Jaligot 2001, Fact 3.1]).

If $H < G$ is a not necessarily full or simple Frobenius group of finite Morley rank and degenerate type, we obtain a weaker version of mock hyperbolic reflection spaces which still allows us to extend Frécon's and Wagner's results:

Theorem 1.4. *If $H < G$ is a connected Frobenius group of Morley rank n and degenerate type with abelian Frobenius complement H of Morley rank k, then $n \geq 2k + 1$.*

If $n = 2k + 1$, then G splits as $G = N \rtimes H$ for some definable connected normal subgroup N. Moreover, if N is solvable, then there is an interpretable field K of characteristic $\neq 2$ such that $G = K_+ \rtimes H$, $H \leq K^*$, and H acts on K_+ by multiplication.

2. Mock hyperbolic reflection spaces

We now introduce the notion of mock hyperbolic reflection spaces, which will be central to our work. The motivating example for our construction comes from sharply 2-transitive groups in characteristic different from 2 (see Section 3) in which the involutions have a rich geometric structure, which is reflected in the following definition.

Let G be a group and $J \subset G$ a conjugacy class of involutions in G, and let $\Lambda \subset \mathcal{P}(J)$ be a G-invariant family of subsets of J such that each $\lambda \in \Lambda$ contains at least two elements. We view involutions in J as points and elements of Λ as lines, so that the conjugation action of J on itself corresponds to point reflections.

For involutions $i \neq j \in J$, we write

$$\ell_{ij} = \{k \in J : ij \in kJ\},$$

and we say that the line ℓ_{ij} *exists in* Λ if $\ell_{ij} \in \Lambda$.

Definition 2.1. Let G be a group, let $J \subset G$ be a conjugacy class of involutions in G, and let $\Lambda \subset \mathcal{P}(J)$ be G-invariant and such that each $\lambda \in \Lambda$ contains at least two elements. The pair (J, Λ) is a *partial mock hyperbolic reflection space* if the following conditions are satisfied:

(a) For all $\lambda \in \Lambda$ and $i \neq j \in \lambda$, we have

$$\lambda = \ell_{ij} = \{k \in J : ij \in kJ\}.$$

In particular, if $i \neq j$ are contained in lines $\lambda, \delta \in \Lambda$, then $\lambda = \ell_{ij} = \delta$. Therefore any two points are contained in at most one line.

(b) Midpoints exist and are unique; i.e., given i, j in J there is a unique $k \in J$ such that $i^k = j$.

(c) Given two distinct lines there is at most one point reflecting one line to the other. In other words, if $\lambda^i = \lambda^j$ for $i \neq j$ in J, then $\lambda^i = \lambda = \lambda^j$.

We say that (J, Λ) is a *mock hyperbolic reflection space* if it satisfies (a)–(c) and furthermore $\ell_{ij} \in \Lambda$ for all $i \neq j \in J$.

Given a group G and a conjugacy class J of involutions in G, in light of Definition 2.1 and in a slight abuse of notation, we say that J forms a mock hyperbolic reflection space if $(J, \{\ell_{ij} : i \neq j \in J\})$ is a mock hyperbolic reflection space.

For a group G and subset $A \subset G$, we write

$$A^{\cdot n} = \{a_1 \cdots a_n \mid a_1, \ldots, a_n \in A\} \subseteq G.$$

Remark 2.2. Let (J, Λ) be a partial mock hyperbolic reflection space. We say that involutions $i, j, k \in J$ are *collinear* if there is some $\lambda \in \Lambda$ with $i, j, k \in \lambda$.

Furthermore, if J is a conjugacy class of involutions in G and $\Lambda \subset \mathcal{P}$ is such that every $\lambda \in \Lambda$ contains at least two elements, we will see below that if (a) and (b) hold, then (c) is equivalent to either of the following conditions:

(c′) If $\lambda^i = \lambda^j$ for $i \neq j$ in J and $\lambda \in \Lambda$, then $i, j \in \lambda$.

(c″) For every line $\lambda \in \Lambda$, we have $N_G(\lambda) \cap J^{\cdot 2} = \lambda^{\cdot 2}$.

If $\Lambda = \{\ell_{ij} : i \neq j \in J\}$, then (a) is equivalent to

$$i, j, k \in J \quad \text{are collinear if and only if} \quad ijk \in J.$$

Example 2.3. Let $\mathbb{H}^n$ be the n-dimensional real hyperbolic space. Then $\mathrm{Isom}(\mathbb{H}^n)$, the group of all isometries of $\mathbb{H}^n$, contains the point-reflections as a conjugacy class J of involutions. J can be identified with $\mathbb{H}^n$, and hence J forms a mock hyperbolic reflection space. In case $n = 2$ the simple group $\mathrm{PSL}_2(\mathbb{R})$ consists of all orientation-preserving isometries of $\mathbb{H}^2$. $\mathrm{PSL}_2(\mathbb{R})$ is generated by the point-reflections and the point-reflections are the only involutions.

Example 2.4. Let A be a uniquely 2-divisible abelian group, and let $\epsilon \in \mathrm{Aut}(A)$ be given by $\epsilon(x) = x^{-1}$. Put $G = A \rtimes \langle \epsilon \rangle$. Then the set of involutions in G is given by $J = A \times \{\epsilon\}$ and J forms a mock hyperbolic reflection space consisting of a single line.

Other examples arise from sharply 2-transitive groups (Section 3) or can be constructed from a class of uniquely 2-divisible Frobenius groups (Section 4).

Lemma 2.5. *Let G be a group and J a conjugacy class of involutions in G such that J acts regularly on itself by conjugation, i.e., J satisfies condition* (b) *in Definition 2.1. Then the following holds for any $i \in J$:*

(a) *iJ is uniquely 2-divisible.*

(b) $J^{\cdot 2} \cap \mathrm{Cen}(i) = \{1\}$.

(c) *$G = (iJ)\,\mathrm{Cen}(i)$ and every $g \in G$ can be written uniquely as $g = ijh$ with $j \in J$ and $h \in \mathrm{Cen}(i)$.*

Proof. (a) Fix $ia \in iJ$. We have to show that there is a unique $b \in J$ such that

$$ia = (ib)^2 = ibib = ii^b.$$

This is exactly condition (b) in Definition 2.1.

(b) Suppose a and b are involutions in J such that $i^{ab} = i$. Then $i^a = i^b$, and hence $a = b$ by the uniqueness in condition (b). Hence $ab = 1$.

(c) Let $g \in G$, and set $k = i^{g^{-1}}$. Then there is a unique $j \in J$ such that $k^{ij} = i$. Now put $h = jig$. Then $g = ijh$ and, we have

$$k^{ij} = i = k^g = k^{ijh} = i^h,$$

and therefore $h \in \mathrm{Cen}(i)$. This shows existence of such a decomposition, and uniqueness follows from part (b). □

In accordance with the terminology from real hyperbolic spaces or from sharply 2-transitive groups, we call elements of the set

$$S = \{\sigma \in J^{\cdot 2} \setminus \{1\} : \ell_\sigma \text{ exists in } \Lambda\} \cup \{1\}$$

translations. Then (b) of Lemma 2.5 implies that nontrivial translations have no fixed points (in their action on J).

B. H. Neumann [1940] showed that a uniquely 2-divisible group admitting a fixed-point-free involutionary automorphism must be abelian. More generally, uniquely 2-divisible groups with involutionary automorphisms can be decomposed as follows:

Proposition 2.6 [Borovik and Nesin 1994, Exercise 14 on p. 73]. *Let G be a uniquely 2-divisible group, and let $\alpha \in \mathrm{Aut}(G)$ be an involutionary automorphism. Define the sets $\mathrm{Inv}(\alpha) = \{g \in G : g^\alpha = g^{-1}\}$ and $\mathrm{Cen}(\alpha) = \{g \in G : g^\alpha = g\}$.*

Then $G = \mathrm{Inv}(\alpha)\,\mathrm{Cen}(\alpha)$, and for every $g \in G$ there are unique $a \in \mathrm{Inv}(\alpha)$ and $b \in \mathrm{Cen}(\alpha)$ such that $g = ab$. In particular, if α has no fixed points, then G is abelian and α acts by inversion.

Lemma 2.7. *Suppose* (J, Λ) *satisfies conditions* (a) *and* (b) *in Definition 2.1. Let* λ *be a line in* Λ.

(a) $N_G(\lambda) \cap J = \lambda$.

(b) *If* $i \in \lambda$, *then* $\lambda^{\cdot 2} = i\lambda$.

(c) *If* a *and* b *are distinct involutions in* J *such that* $ab \in \lambda^{\cdot 2}$, *then* $a, b \in \lambda$.

(d) $\lambda^{\cdot 2}$ *is a uniquely* 2*-divisible abelian group.*

(e) *If* $i, j, k \in J$ *are such that* ℓ_{ij} *and* ℓ_{jk} *exist in* Λ, *then* $\ell_{ij}^{\cdot 2} \cdot \ell_{jk}^{\cdot 2} = \ell_{ij} \cdot \ell_{jk} \subseteq J^{\cdot 2}$.

(f) $N_G(\lambda) = N_G(\lambda^{\cdot 2})$.

Proof. (a) We first show $\lambda \subseteq N_G(\lambda)$: If $\lambda = \ell_{ij}$, then $j^i = iji$, and hence $ij \in j^i J$. Therefore $j^i \in \lambda$.

Now assume $k \in N_G(\lambda) \cap J$ and $\lambda = \ell_{ij}$. We may assume $k \neq i$. Then $i \neq i^k \in \lambda$, and hence $\lambda = \ell_{i^k i}$. Now $i^k i = kk^i$, so $i^k i \in kJ$, and therefore $k \in \lambda$.

(b) Fix $a \neq b$ in λ. Then $ab \in iJ$, and hence $ab = ij$ for some $j \in J$. It remains to show $j \in \lambda$: we have $ab = ij \in Jj = jJ$, and hence $j \in \lambda$.

(c) Suppose $ab = ij$ and $\lambda = \ell_{ij}$. Then $(ij)^a = (ij)^{-1} = ji$, and therefore $\lambda^a = \lambda$, so $a \in N_G(\lambda) \cap J = \lambda$. Now $a\lambda = \lambda^{\cdot 2}$, and hence $b \in \lambda$.

(d) We first show that $\lambda^{\cdot 2} = i\lambda$ is uniquely 2-divisible. Since we know that iJ is uniquely 2-divisible, it remains to show that $i\lambda$ is 2-divisible. Fix $ia \in i\lambda$, say $ia = (ib)^2$ for some $b \in J$. Then $ia = ii^b$, so $a = i^b$, and thus $b \in N_G(\lambda) \cap J = \lambda$.

It remains to show that $\lambda^{\cdot 2} = i\lambda$ is an abelian group. Note that $i\lambda = \lambda i$, and hence $\lambda^{\cdot 2}$ is closed under multiplication and taking inverses. Therefore $\lambda^{\cdot 2}$ is a uniquely 2-divisible group. Moreover, i acts on $\lambda^{\cdot 2}$ as an involutionary automorphism without fixed points. Now Proposition 2.6 implies that $\lambda^{\cdot 2}$ is abelian.

(e) Since j normalizes ℓ_{ij}, we have $\ell_{ij}^{\cdot 2} = \ell_{ij} j$ by (b), and hence the claim follows.

(f) We only need to show that $N_G(\lambda^{\cdot 2}) \subseteq N_G(\lambda)$. Take $g \in N_G(\lambda^{\cdot 2}) \setminus \{1\}$ and fix $i \neq j \in \lambda$. Then $ij \in \lambda^{\cdot 2}$, and hence $i^g j^g \in \lambda^{\cdot 2}$. Therefore $i^g, j^g \in \lambda$ by (c), and thus $\lambda^g = \lambda$. □

Lemma 2.8. *Suppose* (J, Λ) *satisfies* (a) *and* (b) *in Definition 2.1. Then the following are equivalent*:

(a) (J, Λ) *is a partial mock hyperbolic reflection space.*

(b) *Every line* $\lambda \in \Lambda$ *satisfies* $N_G(\lambda) \cap J^{\cdot 2} = \lambda^{\cdot 2}$.

Proof. Suppose that (J, Λ) forms a partial mock hyperbolic reflection space and fix $ij \in N_G(\lambda) \cap J^{\cdot 2}$ and assume $i \neq j \in J$. Then $\lambda^i = \lambda = \lambda^j$, and therefore $i, j \in N_G(\lambda) \cap J = \lambda$. Thus $N_G(\lambda) \cap J^{\cdot 2} = \lambda^{\cdot 2}$.

Conversely, assume $N_G(\lambda) \cap J^{\cdot 2} = \lambda^{\cdot 2}$ and $\lambda^i = \lambda^j$ for $i \neq j \in J$. Then $ij \in \lambda^{\cdot 2}$, and hence $i, j \in \lambda$ by Lemma 2.7(c). This shows $\lambda^i = \lambda = \lambda^j$. □

Proposition 2.9. *Let G be a group and J a conjugacy class of involutions in G, and suppose (J, Λ) is a partial mock hyperbolic reflection space. Then the following holds*:

(a) *If $\lambda = \ell_{ij} \in \Lambda$, then $\lambda^{\cdot 2} = iJ \cap jJ = \mathrm{Cen}(ij) \cap J^{\cdot 2}$.*

(b) *The set $S \setminus \{1\} = \{ij \in J^{\cdot 2} \setminus \{1\} : \ell_{ij}$ exists in $\Lambda\}$ is partitioned by the family $\{\lambda^{\cdot 2} \setminus \{1\} : \lambda \in \Lambda\}$.*

Proof. By Lemma 2.7, (b) follows from (a). In order to prove (a), we first show $\lambda^{\cdot 2} = iJ \cap jJ$. Fix $ia = jb \in iJ \cap jJ$. Then $ab = ij \in \lambda^{\cdot 2}$, and hence $a, b \in \lambda$ by Lemma 2.7(c). This shows $iJ \cap jJ \subseteq \lambda^{\cdot 2}$. Moreover, we have $\lambda^{\cdot 2} = i\lambda = j\lambda$, and hence $\lambda^{\cdot 2} \subseteq iJ \cap jJ$. Thus $\lambda^{\cdot 2} = iJ \cap jJ$.

The group $\lambda^{\cdot 2}$ is abelian and contains ij. Hence $\lambda^{\cdot 2} \subseteq \mathrm{Cen}(ij) \cap J^{\cdot 2}$. Any element $g \in \mathrm{Cen}(ij)$ normalizes $\lambda = \ell_{ij} = \ell_{ji}$. Thus $\mathrm{Cen}(ij) \cap J^{\cdot 2} \subseteq N_G(\lambda) \cap J^{\cdot 2} = \lambda^{\cdot 2}$ (Lemma 2.8), and hence $\mathrm{Cen}(ij) \cap J^{\cdot 2} = \lambda^{\cdot 2}$. □

If i is an involution in J, then we define $\Lambda_i = \{\lambda \in \Lambda : i \in \lambda\}$ to be the set of all lines that contain i.

Proposition 2.10. *Suppose (J, Λ) forms a partial mock hyperbolic reflection space.*

(a) *Suppose $\lambda \cap \lambda^j \neq \varnothing$ for a line λ and an involution j in J. Then $j \in \lambda$, and therefore $\lambda = \lambda^j$.*

(b) *G acts transitively on Λ if and only if $\mathrm{Cen}(i)$ acts transitively on Λ_i for each $i \in J$.*

Proof. (a) Suppose $\{i\} = \lambda \cap \lambda^j$. Then $i = i^j$ and therefore $j = i \in \lambda$.

(b) If $\mathrm{Cen}(i)$ acts transitively on Λ_i, then G is transitive on Λ, because all involutions in J are conjugate.

Now assume G acts transitively on Λ and suppose $i \in \lambda \cap \lambda^g$ for some $g \in G$. By Lemma 2.5, g can be written as $g = ijh$ for some $j \in J$ and $h \in \mathrm{Cen}(i)$. Note that $\lambda^g = \lambda^{ijh} = \lambda^{jh}$, because i is contained in λ.

Since $h \in \mathrm{Cen}(i)$, this implies that i must be contained in λ^j, and hence

$$i \in \lambda \cap \lambda^j.$$

Therefore (a) implies that j must be contained in λ, and hence $\lambda = \lambda^j$. Hence $\lambda^g = \lambda^h$. Since g was arbitrary, this shows that $\mathrm{Cen}(i)$ acts transitively on Λ_i. □

The geometry of a mock hyperbolic reflection space. Recall that a mock hyperbolic reflection space is a partial hyperbolic space such that any two points are contained in a line. As a first step, we show that the geometry of a mock hyperbolic reflection space cannot contain a proper projective plane:

Lemma 2.11. *Suppose that (J, Λ) is a mock hyperbolic reflection space in a group G and that $X \subseteq J$ is a projective plane. That is, suppose*

(a) *for all* $i \neq j \in X$ *the line* ℓ_{ij} *is contained in* X, *and*

(b) *if* λ *and* δ *are lines contained in* X *then* $\lambda \cap \delta \neq \varnothing$.

Then $X^{\cdot 2}$ *is a uniquely* 2*-divisible subgroup of* G.

Proof. The set $X^{\cdot 2}$ is a group by Lemma 2.7(e). This group is uniquely 2-divisible by Lemma 2.7(d) and Proposition 2.9(b). □

Lemma 2.12. *Suppose* J *forms a mock hyperbolic reflection space in a group* G, *and let* $H \subseteq J^{\cdot 2}$ *be a subgroup of* G *which is uniquely* 2*-divisible and normalized by an involution* $i \in J$. *Then* $H \subseteq \mathrm{Cen}(\sigma)$ *for some* $\sigma \in J^{\cdot 2} \setminus \{1\}$.

Proof. Since H is uniquely 2-divisible and i acts as an involutionary automorphism without fixed points, Proposition 2.6 implies that H is abelian and hence must be contained in the centralizer of some translation. □

Proposition 2.13. *If* J *is a mock hyperbolic reflection space in a group* G, *then it does not contain a proper projective plane. That is, if* $X \subseteq J$ *is a projective plane, then* X *contains at most one line.*

Proof. By Lemma 2.11, the set $X^{\cdot 2}$ is a uniquely 2-divisible subgroup of G. By Lemma 2.5(b), each $j \in X$ acts on $X^{\cdot 2}$ as an involutionary automorphism without fixed points. By the previous lemma, $X^{\cdot 2} \leq \mathrm{Cen}(\sigma)$ for some $\sigma \in J^{\cdot 2} \setminus \{1\}$, and hence $X \subseteq \ell_\sigma$ by Lemma 2.7(c). □

Theorem 2.14. *Suppose* J *forms a mock hyperbolic reflection space in a group* G. *Then the following are equivalent*:

(a) Λ *consists of a single line.*

(b) J *is a projective plane.*

(c) G *has an abelian normal subgroup* $A \not\subseteq \bigcap_{i \in J} \mathrm{Cen}(i)$.

(d) $J^{\cdot 2} = iJ$ *for any involution* $i \in J$.

(e) iJ *is commutative for any involution* $i \in J$.

(f) iJ *is a subgroup of* G *for any involution* $i \in J$.

(g) $J^{\cdot 2}$ *is a subgroup of* G.

(h) iJ *is an abelian normal subgroup of* G, *and* G *splits as* $G = iJ \rtimes \mathrm{Cen}(i)$ *for any involution* $i \in J$.

Proof. We show the following implications:

$$\text{(d)} \iff \text{(a)} \implies \text{(b)} \implies \text{(g)} \implies \text{(e)} \implies \text{(f)} \implies \text{(h)} \implies \text{(c)} \implies \text{(a)}.$$

To show (a) $\iff$ (d), assume (d) and fix a line $\lambda = \ell_{ij}$. Then

$$i\lambda = \lambda^{\cdot 2} = iJ \cap jJ = J^{\cdot 2} = iJ,$$

and hence $\lambda = J$ is the only line. Conversely, assume (a) holds and $\lambda = J$ is the unique line. Then $J^{\cdot 2} = \lambda^{\cdot 2} = i\lambda = iJ$ by Lemma 2.7.

(a) $\Longrightarrow$ (b) is trivial.

(b) $\Longrightarrow$ (g) holds by Lemma 2.11.

Now assume (g) holds. By Proposition 2.9, $J^{\cdot 2} \setminus \{1\}$ is partitioned by the family $\{\lambda^{\cdot 2} \setminus \{1\} : \lambda \in \Lambda\}$. Each $\lambda^{\cdot 2}$ is a uniquely 2-divisible abelian group by Lemma 2.7. Therefore $J^{\cdot 2}$ is uniquely 2-divisible. If i is any involution, then i normalizes $J^{\cdot 2}$ and acts by conjugation as an involutionary automorphism without fixed points. Therefore $J^{\cdot 2}$ is an abelian group by Proposition 2.6. In particular, $iJ \subseteq J^{\cdot 2}$ is commutative. This shows (e).

Now assume (e). iJ is partitioned by $\{\lambda^{\cdot 2} : \lambda \in \Lambda, i \in \lambda\}$ and if $\lambda = \ell_{ij}$, then $\lambda^{\cdot 2} = iJ \cap jJ = \mathrm{Cen}(ij) \cap J^{\cdot 2}$ by Proposition 2.9(a). Since iJ is commutative, this implies $iJ = \lambda^{\cdot 2}$, and hence iJ is a subgroup of G by Lemma 2.7. This shows (f).

We next show (f) $\Longrightarrow$ (h): iJ is a uniquely 2-divisible group and i acts as an involutionary automorphism without fixed points. Therefore iJ is an abelian subgroup of G by Proposition 2.6. Note that $N_G(iJ)$ contains $\mathrm{Cen}(i)$ and iJ. Therefore $G = iJ\,\mathrm{Cen}(i) = N_G(iJ)$ by Lemma 2.5. Hence iJ is an abelian normal subgroup of G, and therefore $G = iJ \rtimes \mathrm{Cen}(i)$ by Lemma 2.5.

(h) $\Longrightarrow$ (c) is obvious.

To see that (c) implies (a), let $i \in J$ and $a \in A \setminus \mathrm{Cen}(i)$. Then

$$1 \neq a^{-1}a^i = i^a i \in A \cap iJ.$$

In particular, $A \cap iJ$ is nontrivial. Now fix $\sigma \in (A \cap iJ) \setminus \{1\}$, and set $\lambda = \ell_\sigma$. Then

$$\mathrm{Cen}(\sigma) \cap J^{\cdot 2} = \lambda^{\cdot 2},$$

so $A \cap J^{\cdot 2} \subseteq \lambda^{\cdot 2}$. This implies that $\lambda^{\cdot 2}$ is a normal subset of G, and hence λ is a normal subset of G by Lemma 2.7. Therefore $\lambda = J$ by Lemma 2.7. □

3. Sharply 2-transitive groups

In this section we consider a particular class of Frobenius groups: a permutation group G acting on a set X, where $|X| \geq 2$, is called *sharply* 2*-transitive* if it acts regularly on pairs of distinct points, or equivalently, if G acts transitively on X and for each $x \in X$ the point stabilizer G_x acts regularly on $X \setminus \{x\}$. Thus, a sharply 2-transitive group *splits* if it can be written as a semidirect product of a regular normal subgroup with a point-stabilizer. For two distinct elements $x, y \in X$ the unique $g \in G$ such that $(x, y)^g = (y, x)$ is an involution. Hence the set J of involutions in G is nonempty and forms a conjugacy class.

The (permutation) characteristic of a group G acting sharply 2-transitively on a set X is defined as follows: put $\mathrm{char}(G) = 2$ if and only if involutions have no fixed points. If involutions have a (necessarily unique) fixed point, the G-equivariant bijection $i \mapsto \mathrm{fix}(i)$ allows us to identify the given action of G on X with the conjugation action of G on J. Thus, in this case, the set $S \setminus \{1\}$ of nontrivial translations also forms a single conjugacy class. We put $\mathrm{char}(G) = p$ (or 0) if translations have order p (or infinite order, respectively). For the standard examples of sharply 2-transitive groups, namely $K \rtimes K^*$ for some field K, this definition of characteristic agrees with the characteristic of the field K.

Remark 3.1. Let G be a sharply 2-transitive group of characteristic $\mathrm{char}(G) \neq 2$. Since G acts sharply 2-transitively by conjugation on the set J of involutions in G, the following properties are easy to see:

(a) $\mathrm{Cen}(i)$ acts regularly on $J \setminus \{i\}$.

(b) The set J acts regularly on itself by conjugation, that is, condition (b) of Definition 2.1 holds.

(c) $J^{\cdot 2} \cap \mathrm{Cen}(i) = \{1\}$ for all $i \in J$.

In particular, a nontrivial translation does not have a fixed point.

In order to define the lines for a mock hyperbolic reflection space on J, we need the following equivalent conditions to be satisfied:

Proposition 3.2. *If G is a sharply 2-transitive group of characteristic different from 2, the following conditions are equivalent*:

(a) *Commuting is transitive on* $J^{\cdot 2} \setminus \{1\}$.

(b) $iJ \cap kJ$ *is uniquely* 2*-divisible for all involutions* $i \neq k \in J$.

(c) $\mathrm{Cen}(ik) = iJ \cap kJ$ *is abelian and is inverted by* k *for all* $i \neq k \in J$.

(d) *The set* $\{\mathrm{Cen}(\sigma) \setminus \{1\} : \sigma \in J^{\cdot 2} \setminus \{1\}\}$ *forms a partition of* $J^{\cdot 2} \setminus \{1\}$.

Note that these conditions are satisfied in split sharply 2-transitive groups by Theorem 3.5 whenever $\mathrm{char}(G) = p \neq 0, 2$ or if G satisfies the descending chain condition for centralizers, so in particular if G has finite Morley rank by [Borovik and Nesin 1994, Lemma 11.50].

Proof. For (a) $\Longrightarrow$ (b), note that since $(ij)^2 = ii^j \in iJ$ every element of iJ has a unique square root in iJ. Let $\tau \in iJ \cap kJ$. Since commuting is transitive, the group $A = \langle \mathrm{Cen}(\tau) \cap J^{\cdot 2} \rangle \leq \mathrm{Cen}(\tau)$ is abelian. Moreover, $A \cap J = \varnothing$ by Remark 3.1. Hence the square map is an injective group homomorphism from A to A.

There is $\sigma_i \in iJ$ such that $\sigma_i^2 = \tau$ and so $\sigma_i \in \mathrm{Cen}(\tau) \cap iJ$. Similarly we find $\sigma_k \in \mathrm{Cen}(\tau) \cap kJ$ such that $\sigma_k^2 = \tau$. Since the square map is injective, it follows that $\sigma_i = \sigma_k \in iJ \cap kJ$. Therefore $iJ \cap kJ$ is uniquely 2-divisible.

(b) $\Longrightarrow$ (c) is contained in [Borovik and Nesin 1994, Lemma 11.50(iv)].

(c) $\Longrightarrow$ (d) and (d) $\Longrightarrow$ (a) are obvious. □

The examples constructed in [Rips et al. 2017] (see also [Tent and Ziegler 2016]) show that in characteristic 2 these conditions need not be satisfied. The nonsplit examples in characteristic 0 constructed in [Rips and Tent 2019] satisfy the assumptions and it is an open question whether nonsplit sharply 2-transitive groups exist in characteristic 0 which fail to satisfy these conditions. Note that [Rips and Tent 2019, Lemmas 2.3 and 5.3] imply that the maximal near-field in these examples is planar.

Assume now that the conditions of Proposition 3.2 are satisfied. Then for $i \neq j \in J$ we put

$$\ell_{ij} = \{k \in J : ij \in kJ\} \quad \text{and} \quad \Lambda = \{\ell_{ij} : i \neq j \in J\}.$$

By Remark 3.1(b), (J, Λ) satisfies conditions (a) and (b) of Definition 2.1 and by Proposition 3.2, we have

$$\ell_{ij} = \{k \in J : ij \in kJ\} = i\,\mathrm{Cen}(ij) = \{k \in J : (ij)^k = ji\}.$$

The point-line geometry (J, Λ) is equivalent to the incidence geometry considered by Borovik and Nesin [1994, Section 11.4].

If $\lambda = \ell_{ij}$ is a line, then $N_G(\lambda) = N_G(\mathrm{Cen}(ij))$ is a split sharply 2-transitive group,

$$N_G(\lambda) = \mathrm{Cen}(ij) \rtimes N_{\mathrm{Cen}(i)}(\lambda),$$

and corresponds to the maximal near-field (see, e.g., [Kerby 1974] or [Borovik and Nesin 1994, Chapter 11]). The maximal near-field is called *planar* if

$$N_G(\lambda) = \mathrm{Cen}(ij) \cup \bigcup_{k\in\lambda} N_{\mathrm{Cen}(k)}(\lambda),$$

i.e., if $\mathrm{Cen}(ij)$ coincides with the set of fixed-point-free elements of $N_G(\lambda)$.

Lemma 3.3. *Assume that G is sharply 2-transitive and* $\mathrm{char}(G) \neq 2$, *and if* $\mathrm{char}(G) = 0$, *assume furthermore that G satisfies the descending chain condition on centralizers. Assume moreover that the maximal near-field is planar. If $\lambda \in \Lambda$ and $i \neq j \in J$ such that $\lambda^i = \lambda^j$, then $i, j \in \lambda$ and so $\lambda^i = \lambda = \lambda^j$, and thus condition* (c) *of Definition 2.1 holds.*

Proof. This is contained in the proof of [Borovik and Nesin 1994, Theorem 11.51]. Since our definition of lines is slightly different from the one given in that work, we include a proof. If $\lambda^i = \lambda^j$ then $ij \in N_G(\lambda)$, and hence $ij \in N_G(\lambda^{\cdot 2})$. By Propositions 2.9(a) and 3.2(c), we have $\lambda^{\cdot 2} = \mathrm{Cen}(\sigma)$ for some $\sigma \in J^{\cdot 2} \setminus \{1\}$ such that $\lambda = \ell_\sigma$. Fix $s \in \lambda$. The group $N_G(\mathrm{Cen}(\sigma)) = \mathrm{Cen}(\sigma) \rtimes N_{\mathrm{Cen}(s)}(\mathrm{Cen}(\sigma))$ is split sharply 2-transitive by [Borovik and Nesin 1994, Proposition 11.51]. Since the maximal near-field is planar, we have

$$ij \in N_G(\mathrm{Cen}(\sigma)) \cap J^{\cdot 2} = \mathrm{Cen}(\sigma),$$

and therefore $i, j \in \ell_\sigma = \lambda$. $\square$

Corollary 3.4. *Let G be a sharply 2-transitive group. Then the set of involutions $J \subset G$ forms a mock hyperbolic reflection space in any of the following cases*:

(a) *G is a split sharply 2-transitive group corresponding to a planar near-field of characteristic $\neq 2$;*

(b) $\mathrm{char}(G) = p \neq 0, 2$ *and the maximal near-field is planar*; *or*

(c) $\mathrm{char}(G) = 0$, *G satisfies the descending chain condition for centralizers*, *and the maximal near-field is planar.*

In particular, *if* $\mathrm{char}(G) \neq 2$ *and G is of finite Morley rank*, *then the involutions in G form a mock hyperbolic reflection space.*

In the case of sharply 2-transitive groups, Theorem 2.14 reduces to the following well-known result of Neumann [1940]:

Theorem 3.5. *A sharply 2-transitive group G splits if and only if the set of translations $J^{\cdot 2}$ is a subgroup of G* (*and in that case*, *$J^{\cdot 2}$ must in fact be abelian*).

4. Uniquely 2-divisible Frobenius groups

In this section we will construct (partial) mock hyperbolic reflection spaces from uniquely 2-divisible Frobenius groups with abelian Frobenius complement. This construction makes use of K-loops and quasidirect products.

K-loops and quasidirect products. K-loops are nonassociative generalizations of abelian groups. They are also known as Bruck loops and gyrocommutative gyrogroups. We mostly follow Kiechle's book [2002].

Definition 4.1. A groupoid $(L, \cdot, 1)$ is a *K-loop* if

(a) it is a loop, i.e., the equations
$$ax = b \quad \text{and} \quad xa = b$$
have unique solutions for all $a, b \in L$,

(b) it satisfies the Bol condition, i.e.,
$$a(b \cdot ac) = (a \cdot ba)c$$
for all $a, b, c \in L$, and

(c) it satisfies the automorphic inverse property, i.e., all elements of L have inverses, and we have
$$(ab)^{-1} = a^{-1}b^{-1}$$
for all $a, b \in L$.

Given $a \in L$, let $\lambda_a : L \to L$ be defined by $\lambda_a(x) = ax$. Given $a, b \in L$, we define the precession map
$$\delta_{a,b} = \lambda_{ab}^{-1}\lambda_a\lambda_b.$$

These maps are characterized by

$$a \cdot bx = ab \cdot \delta_{a,b}(x) \quad \text{for all } x \in L.$$

If L is a K-loop, then the precession maps are automorphisms, and we set

$$\mathcal{D} = \mathcal{D}(L) = \langle \delta_{a,b} : a, b \in L \rangle \leq \mathrm{Aut}(L).$$

The following identities will be used in this section:

Proposition 4.2. *Let L be a K-loop, $a, b \in L$, and $\alpha \in \mathrm{Aut}(L)$. Then the following identities hold*:

(a) [Kiechle 2002, 2.4(2)] $\alpha^{-1}\delta_{a,b}\alpha = \delta_{\alpha^{-1}(a),\alpha^{-1}(b)}$.

(b) [Kiechle 2002, 6.1(1)] $\delta_{a,a^{-1}} = \mathrm{id}$.

(c) [Kiechle 2002, Theorem 6.4(1)(VI)] $\delta_{a,ba} = \delta_{a,b}$.

(d) [Kiechle 2002, part of Theorem 3.7] $\delta_{a,b} = \delta_{b,a}^{-1} = \delta_{a^{-1},b^{-1}}$.

Definition 4.3. Let G be a group. A subset $L \subseteq G$ is a twisted subgroup of G if and only if $1 \in L$, $L^{-1} \subseteq L$, and $aLa \subseteq L$ for all $a \in L$.

Note that twisted subgroups are closed under the square map. A twisted subgroup is *uniquely* 2*-divisible* if the square map is bijective.

Proposition 4.4 [Kiechle 2002, Theorem 6.14]. *Let G be a group with a uniquely* 2*-divisible twisted subgroup $L \subseteq G$. Then*

$$a \otimes b = a^{1/2} b a^{1/2}$$

makes L into a K-loop $(L, \otimes, 1)$ and integer powers of elements in L agree in G and $(L, \otimes)$. Given $a, b \in L$, the precession map $\delta_{a,b}$ is given by conjugation with

$$d_{a,b} = b^{1/2} a^{1/2} (a^{1/2} b a^{1/2})^{-1/2}.$$

Proof. The formula for the precession maps follows from simple calculation. Everything else is contained in [Kiechle 2002, Theorem 6.14]. □

Proposition 4.5 [Kiechle 2002, Theorem 2.13]. *Let L be a K-loop, and let $\mathcal{A} \leq$* $\mathrm{Aut}(L)$ *be a group of automorphisms such that $\mathcal{D}(L) \subseteq \mathcal{A}$. Then*:

(a) *The quasidirect product $L \rtimes_Q \mathcal{A}$ given by the set $L \times \mathcal{A}$ together with the multiplication*

$$(a, \alpha)(b, \beta) = (a \cdot \alpha(b), \delta_{a,\alpha(b)}\alpha\beta)$$

forms a group with neutral element $(1, \mathrm{id})$. *Inverses are given by*

$$(a, \alpha)^{-1} = (\alpha^{-1}(a^{-1}), \alpha^{-1}).$$

(b) *$L \rtimes_Q \mathcal{A}$ acts faithfully and transitively on L by*

$$(a, \alpha)(x) = a\alpha(x) \quad \textit{for all } (a, \alpha) \in L \rtimes_Q \mathcal{A} \textit{ and } x \in L.$$

Mock hyperbolic reflection spaces from uniquely 2-divisible Frobenius groups. Let $H < G$ be a uniquely 2-divisible Frobenius group with abelian complement H.

We set L to be the K-loop $L = (G, \otimes)$, where $\otimes$ is defined by

$$a \otimes b = a^{1/2} b a^{1/2}.$$

Set $\mathcal{A} = G \times \langle \epsilon \rangle < \mathrm{Aut}(L)$, where ϵ inverts all elements of L. Put $\mathcal{G} = L \rtimes_Q \mathcal{A}$. Let J be the set of all involutions in $\mathcal{G}$, and put $\iota = (1, \epsilon) \in J$.

Lemma 4.6. (a) $J = L \times \{\epsilon\}$.

(b) $\mathrm{Cen}(\iota) = 1 \times \mathcal{A}$.

(c) *For all* $i, j \in J$, *there is a unique* $k \in J$ *such that* $j = i^k$.

Proof. L is a K-loop by Proposition 4.4.

(a) Fix $(a, \alpha) \in \mathcal{G}$ such that $(a, \alpha)^2 = (1, \mathrm{id})$. Note that

$$(a, \alpha)(a, \alpha) = (a \otimes \alpha(a), \delta_{a,\alpha(a)} \alpha^2).$$

Now $a \otimes \alpha(a) = 1$ implies $\alpha(a) = a^{-1}$, and therefore $\delta_{a,\alpha(a)} = \mathrm{id}$. Hence we must have $\alpha^2 = \mathrm{id}$.

If $\alpha = \mathrm{id}$, then $a \otimes \alpha(a) = a^2$, so $a^2 = 1$, and thus $a = 1$. In that case, $(a, \alpha) = (1, \mathrm{id})$ is the neutral element in $\mathcal{G}$.

This shows $J = L \times \{\epsilon\}$, because ϵ is the only involution in $\mathcal{A}$.

(b) Fix $(a, \alpha) \in \mathrm{Cen}(\iota)$. We have

$$(a, \alpha)(1, \epsilon) = (a, \alpha\epsilon) \quad \text{and} \quad (1, \epsilon)(a, \alpha) = (a^{-1}, \epsilon\alpha).$$

Hence $(a, \alpha) \in \mathrm{Cen}(\iota)$ if and only if $a = a^{-1}$ if and only if $a = 1$.

(c) Take involutions $(a, \epsilon), (b, \epsilon), (c, \epsilon) \in J = L \times \{\epsilon\}$. Then

$$\begin{aligned}(b, \epsilon)(a, \epsilon)(b, \epsilon) &= (b, \epsilon)(a \otimes b^{-1}, \delta_{a,b^{-1}}) \\ &= (b \otimes (a^{-1} \otimes b), \delta_{b,a^{-1}\otimes b} \delta_{a,b^{-1}} \epsilon) \\ &= ((b \otimes a^{-1/2})^2, \epsilon).\end{aligned}$$

Hence we have $(a, \epsilon)^{(b,\epsilon)} = (c, \epsilon)$ if and only if $b \otimes a^{-1/2} = c^{1/2}$. The loop conditions ensure that for all $a, c \in L$ there is a unique b satisfying this equation. □

Now set $\lambda_0 = H \times \{\epsilon\} \subseteq J$, and put $\Lambda = \{\lambda_0^g : g \in \mathcal{G}\}$. We view elements of Λ as lines, and we view involutions as points. Note that Λ is $\mathcal{G}$-invariant and all lines are conjugate.

The following will be shown in this section:

Theorem 4.7. (a) (J, Λ) *is a partial mock hyperbolic reflection space in* $\mathcal{G}$.

(b) *If* G *is full, i.e., if* $G = \bigcup_{g \in G} H^g$, *then* (J, Λ) *is a mock hyperbolic reflection space.*

(c) *Suppose* $i, j, k \in J$ *are pairwise distinct such that the lines* ℓ_{ij} *and* ℓ_{ik} *exist, and assume that* i, j, k *are not collinear. Then* $\mathrm{Cen}_{\mathcal{G}}(i, j, k) = 1$. *In particular,* $\mathcal{G}$ *acts faithfully on* J.

Lemma 4.8. *Let* λ *be a line containing* ι. *Then* λ *is of the form*

$$\lambda = H^g \times \{\epsilon\}$$

for some $g \in G$.

Proof. We have $\lambda = \lambda_0^g$ for some $g = (a, \alpha) \in \mathcal{G}$. So elements of λ are of the form

$$\begin{aligned}(\alpha^{-1}(a^{-1}), \alpha^{-1})(c, \epsilon)(a, \alpha)& \\ &= (\alpha^{-1}(a^{-1}), \alpha^{-1})(c \otimes a^{-1}, \delta_{c,a^{-1}}\epsilon\alpha) \\ &= \big(\alpha^{-1}(a^{-1}) \otimes \alpha^{-1}(c \otimes a^{-1}), \delta_{\alpha^{-1}(a^{-1}),\alpha^{-1}(c\otimes a^{-1})}\alpha^{-1}\delta_{c,a^{-1}}\alpha\epsilon\big) \\ &= \big(\alpha^{-1}(a^{-1} \otimes (c \otimes a^{-1})), \alpha^{-1}\delta_{a^{-1},c\otimes a^{-1}}\delta_{c,a^{-1}}\alpha\epsilon\big)\end{aligned}$$

for some $c \in H$, where the last equality holds by Proposition 4.2(a).

Note that $a^{-1} \otimes (c \otimes a^{-1}) = (a^{-1} \otimes c^{1/2})^2$. We assume $\iota \in \lambda$. Hence

$$1 = a^{-1} \otimes c^{1/2}$$

for some $c \in H$, and thus $a = c^{1/2} \in H$. This implies $\lambda = (\alpha^{-1}(H), \epsilon) \subseteq J$. □

Corollary 4.9. *Any two distinct points are contained in at most one line.*

Lemma 4.10. *Fix distinct involutions* $i, j \in J$ *and suppose* ℓ_{ij} *exists in* Λ. *Then*

$$\begin{aligned}\ell_{ij} &= \{k \in J : ij \in kJ\} \\ &= \{k \in J : (ij)^k = (ij)^{-1}\}.\end{aligned}$$

Proof. We may assume that $\ell_{ij} = H \times \{\epsilon\}$ and $ij = (c, 1)$ for some $c \in H \setminus \{1\}$. The second equality is easy, and therefore we only show the first equality.

We first show $\ell_{ij} \subseteq \{k \in J : ij \in kJ\}$: Take $d \in H \setminus \{1\}$. Then

$$(d, \epsilon)(c, 1)(d, \epsilon) = (d, \epsilon)(c \otimes d, \delta_{c,d}\epsilon) = (d \otimes (c \otimes d)^{-1}, \delta_{d,(c\otimes d)^{-1}}\delta_{c,d}).$$

The Frobenius complement H is abelian, and therefore

$$(d \otimes (c \otimes d)^{-1}, \delta_{d,(c\otimes d)^{-1}}\delta_{c,d}) = (c^{-1}, 1).$$

This shows $(c, 1)^{(d,\epsilon)} = (c, 1)^{-1}$, and hence $\ell_{ij} \subseteq \{k \in J : ij \in kJ\}$.

We now show $\supseteq$ for the first equality: Suppose $(c, 1) = (a, \epsilon)(b, \epsilon)$. We have to show that a is an element of H. We have

$$(c, 1) = (a, \epsilon)(b, \epsilon) = (a \otimes b^{-1}, \delta_{a,b^{-1}}),$$

and hence $a^{1/2}b^{-1}a^{1/2} = a \otimes b^{-1} = c \in H$ and $\delta_{a,b^{-1}} = \mathrm{id}$. By Proposition 4.4, this implies

$$b^{-1/2}a^{1/2}(a \otimes b^{-1})^{-1/2} = 1.$$

So $b^{-1/2}a^{1/2} = c^{1/2}$, and since $c = a^{1/2}b^{-1}a^{1/2}$, this implies $a^{1/2}b^{-1/2} = c^{1/2}$. Hence

$$c^{1/2} = b^{-1/2}a^{1/2} = (a^{1/2}b^{-1/2})^{a^{1/2}} = (c^{1/2})^{a^{1/2}},$$

and therefore $a^{1/2} \in \mathrm{Cen}(c) = H$. □

Lemma 4.11. *Suppose $(a, \alpha) \in N_{\mathcal{G}}(\lambda_0)$. Then $a \in H$ and α normalizes H.*

Proof. Given $c \in H$, we have

$$\begin{aligned}
(a,\alpha)^{-1}(c,\epsilon)(a,\alpha) &= (\alpha^{-1}(a^{-1}),\alpha^{-1})(c\otimes a^{-1},\delta_{c,a^{-1}}\epsilon\alpha)\\
&= \big(\alpha^{-1}(a^{-1})\otimes\alpha^{-1}(c\otimes a^{-1}),\ \delta_{\alpha^{-1}(a^{-1}),\alpha^{-1}(c\otimes a^{-1})}\alpha^{-1}\delta_{c,a^{-1}}\alpha\epsilon\big)\\
&= \big(\alpha^{-1}(a^{-1}\otimes(c\otimes a^{-1})),\ \alpha^{-1}\delta_{a^{-1},c\otimes a^{-1}}\delta_{c,a^{-1}}\alpha\epsilon\big)\\
&= \big(\alpha^{-1}(a^{-1}\otimes(c\otimes a^{-1})),\ \epsilon\big).
\end{aligned}$$

We have $(1, \epsilon) \in \lambda_0$, and therefore

$$1 = a^{-1} \otimes (c_0 \otimes a^{-1})$$

for some $c_0 \in H$. Note that

$$a^{-1} \otimes (c_0 \otimes a^{-1}) = (a^{-1/2}c_0^{1/2}a^{-1/2})^2 = (a^{-1} \otimes c_0^{1/2})^2,$$

and therefore $1 = a^{-1} \otimes c_0^{1/2}$. This shows $a = c_0^{1/2} \in H$.

Moreover, $\alpha^{-1}(a^{-1}\otimes(c\otimes a^{-1})) \in H$ for all $c \in H$, and hence α normalizes H. □

Proposition 4.12. $\qquad N_G(\lambda_0) \cap J^{\cdot 2} = \lambda_0^{\cdot 2}.$

Proof. Fix $a \neq b$ in L such that $(a, \epsilon)(b, \epsilon) = (a \otimes b^{-1}, \delta_{a,b^{-1}}) \in N_{\mathcal{G}}(\lambda_0)$. By Lemma 4.11, we have $a \otimes b^{-1} \in H$ and $\delta_{a,b^{-1}}$ normalizes H.

By Proposition 4.4, the latter is equivalent to

$$b^{-1/2}a^{1/2}(a^{1/2}b^{-1}a^{1/2})^{-1/2} \in H.$$

Since $a^{1/2}b^{-1}a^{1/2} = a \otimes b^{-1} \in H$, this implies $b^{-1/2}a^{1/2} \in H$, and therefore $a^{1/2}b^{-1/2} = a^{1/2}b^{-1}a^{1/2}(b^{-1/2}a^{1/2})^{-1} \in H$.

This shows

$$b^{-1/2}a^{1/2} = (a^{1/2}b^{-1/2})^{b^{1/2}} \in H \cap H^{b^{1/2}}.$$

Thus $b^{1/2} \in h$ and $a^{1/2} = (a^{1/2}b^{-1/2})b^{1/2} \in H$, because H is malnormal in G. □

Proposition 4.13. *Suppose $i, j, k \in J$ are pairwise distinct such that the lines ℓ_{ij} and ℓ_{ik} exist in Λ, and assume that i, j, k are not collinear. Then* $\mathrm{Cen}(i, j, k) = 1$.

Proof. Let $i = (1, \epsilon)$ and fix $j = (a, \epsilon) \in J \setminus \{i\}$. We already know $\mathrm{Cen}(i) = 1 \times \mathcal{A}$. Now fix $(1, \beta) \in \mathrm{Cen}(i) \cap \mathrm{Cen}(j)$. Then

$$(\beta(a), \epsilon\beta) = (1, \beta)(a, \epsilon) = (a, \epsilon)(1, \beta) = (a, \epsilon\beta).$$

Therefore $\beta \in \mathrm{Cen}_{\mathcal{A}}(a) = 1 \times \mathrm{Cen}_G(a)$, and hence $\mathrm{Cen}(i, j) = 1 \times \mathrm{Cen}_G(a)$. This shows the claim, because G is a Frobenius group. □

Proof of Theorem 4.7. We start by checking conditions (a) and (b) of Definition 2.1. Condition (a) follows from Corollary 4.9 and Lemma 4.10. Condition (b) is part (c) of Lemma 4.6.

Now Proposition 4.12 and Lemma 2.8 imply that (J, Λ) is a partial mock hyperbolic reflection space.

If the Frobenius group is full, then it follows from Lemma 4.8 and from the definition of λ_0 that all lines exist and hence that J forms a mock hyperbolic reflection space.

The final statement is Proposition 4.13. □

5. Mock hyperbolic reflection spaces in groups of finite Morley rank

We now turn to the finite Morley rank setting. We refer the reader to [Borovik and Nesin 1994; Poizat 1987] for a general introduction to groups of finite Morley rank. If X is a definable set of finite Morley rank, then we denote its Morley rank by $\mathrm{MR}(X)$ and its Morley degree by $\mathrm{MD}(X)$.

Convention. In the context of finite Morley rank, we say that a definable property P holds for *Morley rank k many elements* if the set defined by P has Morley rank k. In a slight abuse, we may also say that P holds for *generically many* elements of a definable set X if the set of elements in S not satisfying P has smaller Morley rank than X.

We will repeatedly make use of the following:

Proposition 5.1 [Borovik and Nesin 1994, Exercises 11 and 12 on p. 72]. *If G is a group of finite Morley rank and G does not contain an involution, then G is uniquely* 2*-divisible.*

Now let G be a group of finite Morley rank, and let J be a conjugacy class of involutions such that $\mathrm{MD}(J) = 1$. Moreover, we assume that $\Lambda \subseteq \mathcal{P}(J)$ is a G-invariant definable family of subsets of J such that each $\lambda \in \Lambda$ is of the form

$$\lambda = \{k \in J : ij \in kJ\}$$

for any $i \neq j \in \lambda$.

Definition 5.2. We call (J, Λ) a *generic mock hyperbolic reflection space* if (J, Λ) is a partial mock hyperbolic reflection space and for each $i \in J$ the set

$$\{j \in J : \ell_{ij} \in \Lambda\}$$

is generic in J.

Remark 5.3. Let (J, Λ) be a generic mock hyperbolic reflection space.

(a) The condition in the above definition is equivalent to the statement that

$$\{(i, j) \in J^2 : i \neq j \text{ and } \ell_{ij} \text{ exists in } \Lambda\} \subseteq J \times J$$

is a generic subset of J^2.

(b) Write

$$\Lambda(k) = \{\lambda \in \Lambda : \mathrm{MR}(\lambda) = k\}.$$

Fix $i \in J$, and set $B_{(k)}(i) = \{j \in J \setminus \{i\} : \ell_{ij} \in \Lambda(k)\}$. Since $\mathrm{MD}(J) = 1$, there is exactly one $k \leq n$ such that $B_{(k)}(i)$ is a generic subset of J. In that case $(J, \Lambda(k))$ is a generic mock hyperbolic reflection space. Hence we may *assume from now on that all lines in* Λ *have the same Morley rank.*

(c) If (J, Λ) is a generic mock hyperbolic reflection space of finite Morley rank in which all lines have Morley rank k, then we have $\mathrm{MR}(\Lambda) = 2n - 2k$ and $\mathrm{MD}(\Lambda) = 1$ for $n = \mathrm{MR}(J)$. The set of translations

$$S = \{\sigma \in J^{\cdot 2} \setminus \{1\} : \ell_\sigma \text{ exists in } \Lambda\} \cup \{1\}$$

has Morley rank $2n - k$ and Morley degree 1.

If X and Y are definable sets, then we write $X \approx Y$ if X and Y coincide up to a set of smaller rank, i.e., if the sets X, Y, and $X \cap Y$ all have the same Morley rank and Morley degree. This defines an equivalence relation on the family of definable sets. One important property of this equivalence relation is the following:

Proposition 5.4 [Wagner 2017, Lemma 4.3]. *Let G be a group acting definably on a set X in an ω-stable structure. Let Y be a definable subset of X such that $gY \approx Y$ for all $g \in G$. Then there is a G-invariant set $Z \subseteq X$ such that $Z \approx Y$.*

By Theorem 2.14, a mock hyperbolic reflection space consists of one line if and only if the set of translations forms a normal subgroup. For generic mock hyperbolic reflection spaces the following will be shown in this section:

Theorem 5.5. *Suppose (J, Λ) is a generic mock hyperbolic reflection space such that J has Morley rank $\mathrm{MR}(J) = n$. Assume that Λ consists of more than one line and that all lines $\lambda \in \Lambda$ are infinite and of Morley rank $\mathrm{MR}(\lambda) = k$. Then $n \geq 2k + 1$.*

If $n = 2k + 1$, then the translations almost form a normal subgroup: G has a definable connected normal subgroup N of Morley rank $\mathrm{MR}(N) = 2n - k$ such that $N \approx S$. Moreover, $\mathrm{MR}(N \cap \mathrm{Cen}(i)) = n - k$ for any involution $i \in J$.

For the remainder of this section we assume that (J, Λ) is a generic mock hyperbolic reflection space in a group of finite Morley rank G such that (J, Λ) satisfies the assumptions in Theorem 5.5. In particular, $n > k \geq 1$.

Note that we do not state any assumption about the Morley degree of lines.

Generic projective planes.

Definition 5.6. A definable subset $X \subseteq J$ is a *generic projective plane* if

(a) $\mathrm{MR}(X) = 2k$ and $\mathrm{MD}(X) = 1$, and

(b) $\mathrm{MR}(\Lambda_X) = 2k$ and $\mathrm{MD}(\Lambda_X) = 1$,

where Λ_X is the set of all lines $\lambda \subseteq J$ such that $\mathrm{MR}(\lambda \cap X) = k$.

The next lemma follows from easy counting arguments.

Lemma 5.7. *Let $X \subseteq J$ be a definable set of Morley rank $2k$ and Morley degree 1. The following are equivalent*:

(a) *X is a generic projective plane.*

(b) $\mathrm{MR}(\Lambda_X) \geq 2k$.

(c) *The set of $x \in X$ such that* $\mathrm{MR}(\{\lambda \in \Lambda_X : x \in \lambda\}) = k$ *is generic in X.*

Proof. (a) $\Longrightarrow$ (b) This holds by definition.

(b) $\Longrightarrow$ (c) Given $x \in X$ consider $L_x = \{\lambda \in \Lambda_X : x \in \lambda\}$ and note that

$$\mathrm{MR}\Big(\bigcup L_x\Big) = \mathrm{MR}(L_x) + k$$

holds for each $x \in X$. In particular, $\mathrm{MR}(L_x) \leq k$, since $\mathrm{MR}(X) = 2k$. Moreover, $\mathrm{MR}(\Lambda_X) \geq 2k$ and each $\lambda \in \Lambda_X$ is contained in rank k many sets of the form L_x. Hence we must have $\mathrm{MR}(L_x) = k$ for generically many $x \in X$.

(c) $\Longrightarrow$ (b) We have $\mathrm{MR}(X) = 2k$ and $\mathrm{MR}(L_x) = k$ for generically many $x \in X$. Moreover, each $\lambda \in \Lambda_X$ contains rank k many points from X. Thus $\mathrm{MR}(\Lambda_X) \geq 2k$.

(b) $\Longrightarrow$ (a) Consider the set

$$P = \{(x, y) \in X \times X : x \neq y \text{ and } \ell_{xy} \in \Lambda_X\}.$$

Note that each $\lambda \in \Lambda_X$ has rank $2k$ many preimages in P. Since X has rank $2k$ and degree 1, this implies $\mathrm{MR}(\Lambda_X) = 2k$ and $\mathrm{MD}(\Lambda_X) = 1$. □

Lemma 5.8. *Suppose $X \subseteq J$ is a generic projective plane. Then set of $x \in X$ such that $X^x \approx X$ is generic in X.*

Proof. Let $\lambda \in \Lambda_X$ be a line. Recall that $\lambda^{\cdot 2}$ is a group by Lemma 2.7(d). For $i \in \lambda$, set

$$\lambda_i = \{j \in \lambda : ij \in (\lambda^{\cdot 2})^0\} = i(\lambda^{\cdot 2})^0.$$

Then $\{\lambda_i : i \in \lambda\}$ is a partition of λ into sets of rank k and degree 1. Moreover, we have $(\lambda_i)^i = \lambda_i$ for all $i \in \lambda$. In particular, if $\lambda_i \cap X \approx \lambda_i$, then $(\lambda_i \cap X)^i \cap X \approx \lambda_i$.

Hence for all $\lambda \in \Lambda_X$ the set

$$X_\lambda = \{x \in \lambda \cap X : \mathrm{MR}((\lambda \cap X)^x \cap X) = k\}$$

has Morley rank k. Moreover, each $x \in X$ is contained in at most rank k many lines in Λ_X and hence is contained in at most rank k many sets X_λ.

We have $\mathrm{MR}(\Lambda_X) = 2k$, and hence the set

$$\{(x, \lambda) \in X \times \Lambda_X : x \in X_\lambda\}$$

has Morley rank $3k$. Since $\mathrm{MR}(X) = 2k$, this implies that the set of $x \in X$ contained in rank k many sets X_λ is generic in X.

Now if $x \in X_\lambda$ for rank k many λ, then

$$X^x \cap X \supseteq \Bigl(\bigcup_{\lambda : x \in X_\lambda} \lambda \cap X\Bigr)^x \cap X = \bigcup_{\lambda : x \in X_\lambda} (\lambda \cap X)^x \cap X$$

must have Morley rank $2k$, and hence $X^x \approx X$. □

Lemma 5.9. *If $X \subseteq J$ is a generic projective plane and $Z \subseteq J$ is a definable subset with $X \approx Z$, then Z is a generic projective plane.*

Proof. For $x \in X$ put $\Lambda_x = \{\lambda \in \Lambda_X : x \in \lambda\}$. If $\mathrm{MR}(\Lambda_x) = k$, then $B(x) = \bigcup \Lambda_x \approx X$. In particular, $B(x) \approx Z$ for a generic set of $x \in X \cap Z$. If $B(x) \approx Z$, then $\Lambda_x \cap \Lambda_Z$ must have Morley rank k. Hence it follows from Lemma 5.7 that Z must be a generic projective plane. □

Lemma 5.10. *Let $H \leq G$ be a definable subgroup such that* $\mathrm{MR}(H \cap J) = 2k$ *and* $\mathrm{MD}(H \cap J) = 1$. *Then* $\mathrm{MR}(\Lambda_{H\cap J}) < 2k$, *i.e., $H \cap J$ does not form a generic projective plane.*

Proof. This is proved in the same way as [Borovik and Nesin 1994, Proposition 11.71]. Put $Z = H \cap J$.

Assume towards contradiction that $\mathrm{MR}(\Lambda_Z) \geq 2k$. Then Z is a generic projective plane, and hence $\mathrm{MR}(\Lambda_Z) = 2k$ and $\mathrm{MD}(\Lambda_Z) = 1$ (Lemma 5.7).

Let $\lambda \in \Lambda_Z$ be a line. By Proposition 2.10, the family $\{\lambda^i : i \in Z \setminus \lambda\}$ consists of Morley rank $2k$ many lines which do not intersect λ. Therefore the set $\{\delta \in \Lambda_Z : \lambda \cap \delta = \varnothing\} \subseteq \Lambda_Z$ is a generic subset of Λ_Z.

We aim to find a line which intersects Morley rank $2k$ many lines contradicting $\mathrm{MD}(\Lambda) = 1$. For $x \in Z$, set $\Lambda_x = \{\lambda \in \Lambda_Z : x \in \lambda\}$, and set $B(x) = \bigcup \Lambda_x \cap Z \subseteq Z$. Note that $\mathrm{MR}(B(x)) = \mathrm{MR}(\Lambda_x) + k$, and hence $\mathrm{MR}(\Lambda_x) \leq k$ for all $x \in Z$. Since each $\lambda \in \Lambda$ contains Morley rank k many points, we must have $\mathrm{MR}(\Lambda_x) = k$ for a generic set of $x \in Z$.

Fix $x_0 \in Z$ such that Λ_{x_0} has Morley rank $2k$. Then $B(x_0) \subseteq Z$ is generic, and hence $\mathrm{MR}(\Lambda_x) = k$ for a generic set of $x \in B(x_0)$. Since $B(x_0) = \bigcup \Lambda_{x_0}$, we can find a line $\lambda \in \Lambda_{x_0}$ such that $\mathrm{MR}(\Lambda_x) = k$ for a generic set of $x \in \lambda$. But then λ intersects Morley rank $2k$ many lines in Λ_Z. □

Proposition 5.11. *J does not contain a generic projective plane X.*

Proof. Assume $X \subseteq J$ is a generic projective plane, and put

$$H = N_G^{\approx}(X) = \{g \in G : X^g \approx X\}.$$

By Lemma 5.8, the set $X \cap H$ is generic in X. Hence $\mathrm{MR}(H \cap J) \geq 2k$.

Now consider the action of G on J by conjugation. Note that, by Proposition 5.4, there is a definable subset $Z \subseteq J$, $X \approx Z$, such that H normalizes Z. Since J forms a generic mock hyperbolic space, J acts regularly on itself, and hence $\mathrm{MR}(H \cap J) \leq \mathrm{MR}(Z) = 2k$. Therefore $\mathrm{MR}(H \cap J) = 2k$ and $\mathrm{MD}(H \cap J) = 1$ (since $\mathrm{MD}(Z) = 1$). This contradicts Lemma 5.10. □

A rank inequality and a normal subgroup. A line $\lambda \in \Lambda$ is called *complete for some $i \in J \setminus \lambda$* if the set $\{j \in \lambda : \ell_{ij} \in \Lambda\}$ is a generic subset of λ.

Definition 5.12. Let (i, j, p) be a triple of noncollinear involutions in J.

- (i, j, p) is *good* if ℓ_{ij}, ℓ_{jp} exist and ℓ_{ij} is complete for p.
- (i, j, p) is *perfect* if ℓ_{ij}, ℓ_{jp} exist and

$$\{j' \in \ell_{jp} : \ell_{ij'} \in \Lambda \text{ is complete for } p' = j'jp\}$$

  is generic in ℓ_{jp}.

Lemma 5.13. *A generic triple $(i, j, p) \in J^3$ is good. In particular, for any $i \in J$ a generic element of $\{i\} \times J^2$ is good.*

Proof. Fix $i \in J$, and put $B(i) = \{j \in J : \ell_{ij} \in \Lambda\}$. Then $B(i)$ is a generic subset J. Now fix $p \in J \setminus \{i\}$. We aim to show that (i, j, p) must be good for generically many $j \in J \setminus \{i, p\}$.

Note that $B(i)$ and $B(p)$ are generic subsets of J. Therefore $B(i) \cap B(p)$ must be generic in $B(i)$ and $B(i) \setminus B(p)$ is not generic in $B(i)$. Note that

$$B(i) \cap B(p) = \bigcup_{\lambda \in \Lambda_i} (\lambda \cap B(p)) \quad \text{and} \quad B(i) \setminus B(p) = \bigcup_{\lambda \in \Lambda_i} (\lambda \setminus B(p)).$$

Since $\mathrm{MR}(J) = \mathrm{MR}(\Lambda_i) + p$, the set

$$\{\lambda \in \Lambda_i : \mathrm{MR}(\lambda \setminus B(p)) < p\}$$

must be generic in Λ_i.

Hence $\lambda \cap B(p) \approx \lambda$ for generically many $\lambda \in \Lambda_i$. Moreover, if $\lambda \cap B(p) \approx \lambda$ for some $\lambda \in \Lambda_i$ and j is contained in $\lambda \setminus \{i, p\}$, then (i, j, p) is good. The last sentence follows since all elements in J are conjugate. □

Proposition 5.14. *A generic triple $(i, j, p) \in J^3$ is perfect, and for any $i \in J$ a generic element of $\{i\} \times J^2$ is perfect.*

Proof. Since J is a generic mock hyperbolic reflection space, the set $U = \{(j, p) : jp \in S \setminus \{1\}\} \subseteq J^2$ is generic in J^2. For $\sigma \in S$ put $U_\sigma = \{(j, p) : jp = \sigma\}$. Then each U_σ has Morley rank p and U is the disjoint union

$$U = \bigcup_{\sigma \in S} U_\sigma \subseteq J \times J.$$

Now fix $i \in J$. A generic triple in $\{i\} \times U$ is good, and we have $\mathrm{MD}(\{i\} \times U) = 1$. Since $\mathrm{MD}(S) = 1$, this implies that for generically many $\sigma \in S$ the set

$$\{(i, r, s) \in \{i\} \times U_\sigma : (i, r, s) \text{ is good}\}$$

is a generic subset of $\{i\} \times U_\sigma$.

Moreover, if a generic triple in $\{i\} \times U_\sigma$ is good, then a generic triple in $\{i\} \times U_\sigma$ must be perfect. This proves the lemma. □

Now let $\mu : J^3 \to G$ be the multiplication map, and put

$$T = \{(i, j, p) \in J^3 : \ell_{jp} \text{ exists}\} \quad \text{and} \quad T_{\text{perf}} = \{(i, j, p) \in J^3 : (i, j, p) \text{ is perfect}\}.$$

Note that $T_{\text{perf}} \subseteq T$. If (J, Λ) is a mock hyperbolic reflection space, i.e., if all lines exist, then T_{perf} consists of all triples of noncollinear involutions in J.

Lemma 5.15. $$\mathrm{MR}(\mu(T_{\text{perf}})) \geq 2n - k.$$

Proof. For any $i \in J$ the set $\{(j, p) \in J^2 : (i, j, p) \text{ is perfect}\}$ has Morley rank $2n$ by Proposition 5.14. Clearly $ijp = ij'p'$ if and only if $jp = j'p'$. If ℓ_{jp} exists, the set $\{(j', p') \in J^2 : jp = j'p'\}$ has Morley rank k. Hence $\mu(T_{\text{perf}})$ has Morley rank at least $2n - k$. □

Proposition 5.16. *Suppose* $\mathrm{MR}(\mu(T_{\text{perf}})) = 2n - k$. *Then G has a definable connected normal subgroup N of Morley rank* $\mathrm{MR}(N) = 2n - k$ *such that $N \approx S$. Moreover,* $\mathrm{MR}(N \cap \mathrm{Cen}(i)) = n - k$ *for any involution $i \in J$.*

Proof. Set $d = \mathrm{MD}(\mu(T_{\text{perf}}))$ and write $\mu(T_{\text{perf}})$ as a disjoint union

$$\mu(T_{\text{perf}}) = Y_1 \cup \cdots \cup Y_d,$$

where each Y_r has rank $2n - k$ and degree 1. Put $T_i = T_{\text{perf}} \cap (\{i\} \times J \times J)$. Then each T_i has rank $2n$ and degree 1 by Proposition 5.14. Moreover, $\mu(T_i)$ has rank $2n-k$ and degree 1. We can find $1 \le f \le d$ such that

$$\mu(T_i) \approx Y_f$$

for generically many $i \in J$. Put $Y = Y_f$, set $N = \mathrm{Stab}^{\approx}(Y) = \{g \in G : gY \approx Y\}$, and note that N must be a normal subgroup of G, because Y is G-normal up to $\approx$-equivalence.

Now, by Proposition 5.4, there is some $Z \approx Y$ such that $N \subseteq \mathrm{Stab}(Z)$. In particular, N has rank $\le 2n-k$, since $\mathrm{MR}(Z) = 2n-k$. Moreover, if $\mathrm{MR}(N) = 2n-k$, then we must have $N = \mathrm{Stab}(Z)$, since $\mathrm{MD}(Z) = 1$.

Let $J_Y = \{i \in J : \mu(T_i) \approx Y\}$. Given $i \ne j \in J_Y$, we have

$$ij\mu(T_j) \approx \mu(T_i),$$

and hence $ij \in N$. Therefore $J_Y^{\cdot 2} \subseteq N$. Since J_Y is a generic subset of J, we have $\Lambda_{J_Y} \approx \Lambda$, and therefore $J_Y^{\cdot 2} \cap S \approx S$. Thus $\mathrm{MR}(N) = 2n-k$, and hence $N = \mathrm{Stab}(Z)$ is connected. In particular, $J_Y^{\cdot 2} \cap S$ is generic in both N and S, and hence $N \approx S$.

We now show $\mathrm{MR}(N \cap \mathrm{Cen}(i)) = n-k$ for any involution $i \in J$: Fix an involution $i \in J$. If $i \in N$, then $iJ \subseteq N$, and hence $N = iJ(N \cap \mathrm{Cen}(i))$ by Lemma 2.5, and therefore $\mathrm{MR}(N \cap \mathrm{Cen}(i)) = n-k$.

If $i \notin N$, then note that $iJ \cap N$ must be a generic subset of iJ, and therefore the conjugacy class i^N is generic in J. This implies that $N \rtimes \langle i \rangle$ must contain J, and hence is a normal subgroup of G. Now argue as in the first case. □

For $\alpha \in \mu(T)$, we set

$$X_\alpha = \{i \in J : \exists (j, p) \in J \times J \text{ such that } (i, j, p) \in T \text{ and } ijp = \alpha\}.$$

Note that $\mathrm{MR}(\mu^{-1}(\alpha) \cap T) = \mathrm{MR}(X_\alpha) + k$.

If A and B are definable sets, then we write $A \subsetapprox B$ if A is almost contained in B, i.e., if $A \cap B \approx A$.

Lemma 5.17. *Fix a triple $(i, j, p) \in T$.*

(a) *If (i, j, p) is good, then $\ell_{ij} \subsetapprox X_{ijp}$.*

(b) *If (i, j, p) is perfect, then $\ell_{it} \subsetapprox X_{ijp}$ for generically many $t \in \ell_{jp}$. In particular, $\mathrm{MR}(X_{ijp}) \ge 2k$.*

Proof. (a) Since (i, j, p) is good, the line ℓ_{ij} is p-complete. Hence $\ell_{j'p}$ exists for generically many $j' \in \ell_{ij}$. Fix such an j' and write $ij = i'j'$. Then (i', j', p) is good, and hence $i' \in X_{ijp}$.

(b) This follows immediately from (a). □

Lemma 5.18. *Set $l = \mathrm{MR}(\mu(T_{\mathrm{perf}})) - (2n-k)$. Then $2k \le \mathrm{MR}(X_\alpha) \le n-l$ for generically many $\alpha \in \mu(T_{\mathrm{perf}})$. In particular, $n \ge 2k+l$.*

Proof. We have that $\mathrm{MR}(\mu^{-1}(\alpha)\cap T) = \mathrm{MR}(X_\alpha)+k$ for each $\alpha \in \mu(T)$ and that T has rank $3n$, and we trivially have

$$\bigcup_{\alpha\in\mu(T_{\mathrm{perf}})} \mu^{-1}(\alpha)\cap T \subseteq T.$$

Therefore a generic $\alpha \in \mu(T_{\mathrm{perf}})$ must satisfy the inequality

$$\mathrm{MR}(\mu(T_{\mathrm{perf}})) + \mathrm{MR}(X_\alpha) + k \leq \mathrm{MR}(T) = 3n.$$

Moreover, we have $\mathrm{MR}(X_\alpha) \geq 2k$ by Lemma 5.17. Hence

$$2k \leq \mathrm{MR}(X_\alpha) \leq \mathrm{MR}(T) - k - \mathrm{MR}(\mu(T_{\mathrm{perf}})) = n - l$$

for generically many $\alpha \in \mu(T_{\mathrm{perf}})$. □

Proposition 5.19. *Set $l = \mathrm{MR}(\mu(T_{\mathrm{perf}})) - (2n-k)$. Then $n > 2k+l$. In particular, $n > 2k$.*

Proof. Assume not. Then $n = 2k + l$ and $\mathrm{MR}(X_\alpha) = 2k$ for generically many $\alpha \in \mu(T_{\mathrm{perf}})$. Set $M = \{\alpha \in \mu(T_{\mathrm{perf}}) : \mathrm{MR}(X_\alpha) = 2k\}$. This is a generic subset of $\mu(T_{\mathrm{perf}})$. We have

$$\mathrm{MR}\Bigg(\bigcup_{\alpha\in M} \mu^{-1}(\alpha)\cap T\Bigg) = (2n-k+l) + 3k = 6k + 3l = 3n.$$

So $\bigcup_{\alpha\in M}\mu^{-1}(\alpha)\cap T$ is a generic subset of $J \times J \times J$. Note $\mathrm{MR}(M) = 3k+3l$. Therefore we can find $\alpha \in M$ such that $\mu^{-1}(\alpha)\cap T$ has rank $3k$ and contains rank $3k$ many perfect triples. Set $X = X_\alpha$ and $\Lambda_X = \{\lambda \in \Lambda : \lambda \precsim X\}$. Now Lemma 5.17 implies that for a generic $i \in X$ the set

$$\{\lambda \in \Lambda_X : i \in \lambda\}$$

has Morley rank k. Hence $\mathrm{MR}(\Lambda_X) = 2k$, and therefore a degree 1 component of X must be a generic projective plane. This contradicts Proposition 5.11. □

Proof of Theorem 5.5. Set $l = \mathrm{MR}(\mu(T_{\mathrm{perf}})) - (2n-k)$. By Proposition 5.19, we have $2k+1 = n > 2k+l$, and hence $l = 0$. Now Proposition 5.16 implies the theorem. □

6. Frobenius groups of finite Morley rank

We now consider Frobenius groups of finite Morley rank. If G is a group of finite Morley rank and H is a Frobenius complement in G, then H is definable by [Borovik and Nesin 1994, Proposition 11.19]. If G splits as $G = N \rtimes H$, then N is also definable by [Borovik and Nesin 1994, Proposition 11.23].

Epstein and Nesin showed that if $H < G$ is a Frobenius group of finite Morley rank and H is finite, then $H < G$ splits [Borovik and Nesin 1994, Theorem 11.25]. As a consequence it suffices to consider connected Frobenius groups of finite Morley rank [Borovik and Nesin 1994, Corollary 11.27].

Solvable Frobenius groups of finite Morley rank split and their structure is well understood [Borovik and Nesin 1994, Theorem 11.32].

Lemma 6.1. *Let $H < G$ be a connected Frobenius group of finite Morley rank with Frobenius complement H, and let $X \subseteq H \setminus \{1\}$ be a definable H-normal subset such that* $\mathrm{MR}(X) = \mathrm{MR}(H)$. *Then $\bigcup_{b \in G} X^b \subseteq G$ is a generic subset of G.*

Proof. Set $n = \mathrm{MR}(G)$ and $k = \mathrm{MR}(H)$. Consider the map $\alpha : G \times X \to G$, $(b, x) \mapsto x^b$. If $x^b = y^c$ for $x, y \in X$ and $b, c \in G$, then bc^{-1} must be contained in $N_G(H) = H$. Therefore we have

$$\alpha^{-1}(x^b) = \{(c, x^{bc^{-1}}) \in G \times X : bc^{-1} \in H\}.$$

Hence all fibers of α have Morley rank k. This shows that $\alpha(G \times X) = \bigcup_{b \in G} X^b$ must have Morley rank n, and hence is a generic subset of G. $\square$

Groups of finite Morley rank can be classified by the structure of their 2-Sylow subgroups. In case of Frobenius groups this classification is simpler:

Proposition 6.2. *Let G be a connected Frobenius group of finite Morley rank with Frobenius complement H. Then H is connected and G lies in one of the following mutually exclusive cases*:

(a) *H contains a unique involution, and G is of odd type*;

(b) *G does not contain any involutions, and in particular, G is of degenerate type*;

(c) *$G \setminus \left(\bigcup_{g \in G} H^g\right)$ contains involutions, and G is of even type.*

Proof. We first show that H must be connected: if H is not connected, then $\bigcup_{g \in G}(H^0 \setminus \{1\})^g$ and $\bigcup_{g \in G}(H \setminus H^0)^g$ would be two disjoint generic subsets of G. This is impossible, because G is connected.

If H contains an involution, then Delahan and Nesin showed this involution must be unique and moreover all involutions in G are conjugate, so $G \setminus \left(\bigcup_{g \in G} H^g\right)$ cannot contain any involution [Borovik and Nesin 1994, Lemma 11.20]. In particular, G is of odd type, because the connected subgroup H contains a unique involution.

If $G \setminus \left(\bigcup_{g \in G} C^g\right)$ contains an involution, then the proof of [Altınel et al. 2019, Theorem 2] shows that G is of even type. $\square$

Remark 6.3. If $H < G$ is of even type, then Altınel, Berkman, and Wagner showed in [Altınel et al. 2019] that there is a definable normal subgroup N such that $N \cap H = 1$ and N contains all involutions of G. By [Borovik and Nesin 1994, Lemma 11.38], either $G = N \rtimes H$ splits or $HN/N < G/N$ is a Frobenius group of finite Morley rank. Now if $HN/N < G/N$ splits, then it is easy to see that $H < G$ must split. Hence a nonsplit Frobenius group of minimal Morley rank cannot be of even type. Therefore to show that all Frobenius groups of finite Morley rank split, it suffices to consider Frobenius groups of odd and degenerate type.

Frobenius groups of odd type. Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type. Note that G contains a single conjugacy class of involutions, which we denote by J. Moreover, J has Morley degree 1.

Proposition 6.4 [Borovik and Nesin 1994, Proposition 11.18]. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type and J its set of involutions. If $a \in J^{\cdot 2} \setminus \{1\}$ and $i \in J$, then* $\mathrm{Cen}(a) \cap \mathrm{Cen}(i) = \{1\}$.

Lemma 6.5. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type and J its set of involutions. Fix distinct involutions $i, j \in J$.*

(a) *If $a \in iJ \setminus \{1\}$, then $\mathrm{Cen}(a) \subseteq iJ$ is a uniquely 2-divisible abelian group.*

(b) *iJ is uniquely 2-divisible.*

(c) *J acts regularly on itself; i.e., given $i, j \in J$ there is a unique $p \in J$ such that $j = i^p$.*

(d) *$iJ \cap jJ$ is uniquely 2-divisible.*

(e) $\mathrm{Cen}(ij) = iJ \cap jJ$.

(f) *The family $\{\mathrm{Cen}(a) \setminus \{1\} : a \in J^{\cdot 2} \setminus \{1\}\}$ forms a partition of $J^{\cdot 2} \setminus \{1\}$.*

Proof. (a) By the previous proposition, we have $\mathrm{Cen}(a) \cap \mathrm{Cen}(k) = \{1\}$ for all involutions k. In particular, $\mathrm{Cen}(a)$ does not contain an involution and hence is uniquely 2-divisible by Proposition 5.1. Note that i acts on $\mathrm{Cen}(a)$ as a fixed-point-free involutionary automorphism. Hence, by Proposition 2.6, $\mathrm{Cen}(a)$ is abelian and inverted by i, therefore $i\,\mathrm{Cen}(a) \subseteq J$, and we have $\mathrm{Cen}(a) \subseteq iJ$.

(b) Fix $a = ip \in iJ \setminus \{1\}$. Since $\mathrm{Cen}(a) \subseteq iJ$ is uniquely 2-divisible, we have $a = b^2$ for some $b = iq \in iJ$. If $a = c^2$ for another element $c = ir \in iJ$, then $ii^q = ii^r$, and hence $qr \in \mathrm{Cen}(i) \cap J^{\cdot 2} = \{1\}$. Thus $b = c$.

(c) Note that $j = i^p$ if and only if $ij = ii^p = (ip)^2$. Since iJ is uniquely 2-divisible, p exists and is unique.

(d) It suffices to show that $iJ \cap jJ$ is 2-divisible. Given $a \in iJ \cap jJ$, we have $\mathrm{Cen}(a) \subseteq iJ \cap jJ$, and hence $a = b^2$ for some $b \in \mathrm{Cen}(a) \subseteq iJ \cap jJ$.

(e) By (a), we have $\mathrm{Cen}(ij) \subseteq iJ \cap jJ$. Hence it remains to show that $iJ \cap jJ \subseteq \mathrm{Cen}(ij)$. Given $a \in iJ \cap jJ$, a is inverted by i and j, and hence $a \in \mathrm{Cen}(ij)$.

(f) Suppose $c \in \mathrm{Cen}(a) \cap \mathrm{Cen}(b)$ for some $c \neq 1$. Then $a, b \in \mathrm{Cen}(c)$, and hence $a \in \mathrm{Cen}(b)$, because $\mathrm{Cen}(c)$ is abelian. This implies $\mathrm{Cen}(b) \subseteq \mathrm{Cen}(a)$, because $\mathrm{Cen}(b)$ is abelian. Hence $\mathrm{Cen}(a) = \mathrm{Cen}(b)$ by symmetry. This implies (f). □

Given two distinct involutions $i \neq j$ in J, we define the line

$$\ell_{ij} = \{p \in J : (ij)^p = (ij)^{-1}\}.$$

Lemma 6.6. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type and J its set of involutions. Let $i \neq j \in J$. Then $i\ell_{ij} = \mathrm{Cen}(ij)$.*

Proof. Clearly $i\ell_{ij} \subseteq \mathrm{Cen}(ij)$. On the other hand, we have $\mathrm{Cen}(ij) \subseteq iJ$ by Lemma 6.5(e). Given $\sigma = ip \in \mathrm{Cen}(ij)$, we have

$$(ji)^p = (ij)^{ip} = (ij)^\sigma = ij.$$

Therefore $(ij)^p = (ij)^{-1}$, and thus $p \in \ell_{ij}$. □

Lemma 6.7. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type and J its set of involutions. Fix $i \neq j \in J$, and let $p, q \in \ell_{ij}$ be distinct involutions. Then $\ell_{pq} = \ell_{ij}$.*

Proof. We have $pq \in \mathrm{Cen}(ij)$, and hence $\mathrm{Cen}(pq) = \mathrm{Cen}(ij)$. Moreover, $ip \in \mathrm{Cen}(ij) = \mathrm{Cen}(pq)$, and hence

$$\ell_{pq} = p\,\mathrm{Cen}(pq) = i\,\mathrm{Cen}(ij) = \ell_{ij}.$$ □

Hence the set J together with the above notion of lines satisfies conditions (a) and (b) of Definition 2.1.

Lemma 6.8. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type and J its set of involutions. Let $i, j \in J$ be distinct involutions, and let T be a subgroup of G such that $\mathrm{Cen}(ij) \leq T \leq N_G(\mathrm{Cen}(ij))$. Then T can be written as a semidirect product $T = \mathrm{Cen}(ij)(T \cap \mathrm{Cen}(i))$.*

Proof. Note G can be decomposed as $G = iJ\,\mathrm{Cen}(i)$, and put $\lambda = \ell_{ij}$. Given $t \in T$, we can write $t = ipg$ for (unique) elements $k \in J$ and $g \in \mathrm{Cen}(i)$. Then

$$\lambda = \lambda^t = \lambda^{pg}.$$

In particular, $i \in \lambda^p \cap \lambda$. If $\lambda^p \cap \lambda = \{i\}$, then $p = i \in \lambda$. If $\lambda^p = \lambda$, then $p \in \lambda$ by part (a) of Lemma 2.7. Therefore $t = ipg \in \mathrm{Cen}(ij)(T \cap \mathrm{Cen}(i))$. □

We will make use of the following result about conjugacy of complements:

Proposition 6.9 [Borovik and Nesin 1994, Theorem 9.11]. *Let G be a group of finite Morley rank and $H \triangleleft G$ be a definable normal nilpotent subgroup. Assume that G/H is abelian and* $\mathrm{Cen}(g) = 1$ *for some $g \in G$. Then $G = H \rtimes \mathrm{Cen}(g)$ and any two complements of H in G are H-conjugate. Furthermore, $[H, g] = H$.*

Proposition 6.10. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type and J its set of involutions. If $i \neq j$ are two distinct involutions in J, then*

$$N_G(\mathrm{Cen}(ij)) \cap J^{\cdot 2} = \mathrm{Cen}(ij).$$

Proof. Assume there is $a \in \left(N_G(\mathrm{Cen}(ij)) \cap J^{\cdot 2}\right) \setminus \mathrm{Cen}(ij)$ and consider the group $A = \mathrm{Cen}(a) \cap N_G(\mathrm{Cen}(ij))$. Note that A and $\mathrm{Cen}(ij)$ are abelian by Lemma 6.5, and by Lemma 6.5(f), we obtain a semidirect product $K = \mathrm{Cen}(ij) \rtimes A$. Moreover, K is a solvable subgroup of $N_G(\mathrm{Cen}(ij))$. By Lemma 6.8, we have $K = \mathrm{Cen}(ij) \rtimes (K \cap \mathrm{Cen}(i))$. Now Proposition 6.9 implies that A and $K \cap \mathrm{Cen}(i)$ are conjugate. This is impossible by Proposition 6.4. □

Theorem 6.11. *Let $H < G$ be a connected Frobenius group of Morley rank n and odd type, and let J be the set of all involutions in G.*

(a) *J forms a mock hyperbolic reflection space, and all lines in J are infinite.*

(b) *Choose Λ such that (J, Λ) is a generic mock hyperbolic reflection space such that all lines are of Morley rank k, and set $n = \mathrm{MR}(J)$. If $n \leq 2k+1$, then G splits.*

Proof. (a) We first show that J forms a mock hyperbolic reflection space. We already know that conditions (a) and (b) of Definition 2.1 are satisfied. Fix a line $\lambda = \ell_{ij}$. Then $\lambda^{\cdot 2} = i\lambda$ by Lemma 2.7, and hence $\lambda^{\cdot 2} = \mathrm{Cen}(ij)$ by Lemma 6.6. Therefore J forms a mock hyperbolic reflection space by Lemma 2.8.

Moreover, by [Borovik et al. 2007, Proposition 1.1], the centralizer of any element in a connected nontrivial group of finite Morley rank is infinite. In particular, $\mathrm{Cen}(ij)$ is infinite, and therefore all lines in J must be infinite.

(b) Note that if the mock hyperbolic reflection space J consists of a single line, then $H < G$ splits by Theorem 2.14. Hence, by Theorem 5.5, we may assume $n = 2k+1$. Then again by Theorem 5.5, B has a connected normal subgroup N of rank $2n-k$ such that $N \approx S$, where

$$S = \{\sigma \in J^{\cdot 2} \setminus \{1\} : \ell_\sigma \text{ exists}\} \cup \{1\}$$

is the set of translations. Recall that $\mathrm{MR}(S) = 2n-k$ and $\mathrm{MD}(S) = 1$.

On the other hand $N \cap \mathrm{Cen}(i) < N$ is a connected Frobenius group of finite Morley rank, and hence $\bigcup_{i \in J} N \cap \mathrm{Cen}(i) \subseteq N$ is a generic subset of N. This contradicts $N \approx S$. □

As a direct consequence, we get the following known corollary (which also follows from [Borovik and Nesin 1994, Lemma 11.21 and Theorem 11.32]).

Corollary 6.12. *Let $H < G$ be a connected Frobenius group of finite Morley rank of odd type. If G has a nontrivial abelian normal subgroup, then G splits.*

Proof. This follows directly from Theorem 2.14. □

Proposition 6.13. *Let $H < G$ be a connected nonsplit Frobenius group of finite Morley rank and odd type, and let (J, Λ) be the associated mock hyperbolic reflection space. If generic lines have Morley rank* 1, *then G is a nonsplit sharply 2-transitive group of characteristic $\neq 2$.*

Proof. Set $n = \mathrm{MR}(J)$. The set of translations has Morley rank $2n - 1$ and is not generic in G. On the other hand, $\mathrm{Cen}(i)$ acts on $J \setminus \{i\}$ without fixed points. Therefore $\mathrm{MR}(\mathrm{Cen}(i)) \leq 2n$. Hence $G = iJ\,\mathrm{Cen}(i)$ must have Morley rank $2n$ and $\mathrm{Cen}(i)$ has Morley rank n. This implies that $\mathrm{Cen}(i)$ acts regularly on $J \setminus \{1\}$, and hence G is a sharply 2-transitive group. □

Remark 6.14. We will see in Corollary 7.5 that the group in the above proposition must in fact be simple.

Proposition 6.15. *Let $H < G$ be a connected Frobenius group of Morley rank at most* 10 *and odd type. Then either $H < G$ splits or G is a simple nonsplit sharply* 2*-transitive group of Morley rank* 8 *or* 10.

Proof. Assume G does not split. Suppose the set of involutions has Morley rank n and the lines in the associated generic mock hyperbolic reflection space have rank k. Since the set of translations is not generic in G, we have $\mathrm{MR}(G) > 2n-k$. Moreover, we know $n > 2k+1$ and $k \geq 1$. This shows $\mathrm{MR}(G) > 2(2k+2)-k = 3k+4$. Since $\mathrm{MR}(G) \leq 10$, we obtain $k = 1$ and $\mathrm{MR}(G) > 7$. The previous proposition and the remark show that G is a simple sharply 2-transitive group, and hence $\mathrm{MR}(G)$ must be an even number, so $\mathrm{MR}(G)$ is either 8 or 10. □

Frobenius groups of odd type with nilpotent complement. Delahan and Nesin showed that a sharply 2-transitive group of finite Morley rank of characteristic $\neq 2$ with nilpotent point stabilizer must split [Borovik and Nesin 1994, Theorem 11.73]. We will show that the same is true for a Frobenius group of odd type if the lines in the associated mock hyperbolic reflection geometry are strongly minimal or if there is no interpretable bad field of characteristic 0.

We fix a connected Frobenius group $H < G$ of finite Morley rank of odd type, and we denote the set of involutions by J. By Theorem 6.11, J forms a mock hyperbolic reflection space with infinite lines. Note that $H = \mathrm{Cen}(i)$ if i is the unique involution in H. If λ is a line containing i, then $N_G(\lambda) = \lambda^{\cdot 2} \rtimes N_H(\lambda)$ is a split Frobenius group by Theorem 2.14.

Lemma 6.16. *If i and j are involutions with $i \neq j$, then $\mathrm{Cen}(ij)$ has infinite index in $N_G(\mathrm{Cen}(ij))$. In particular, $N_{\mathrm{Cen}(i)}(\ell_{ij})$ is infinite.*

Proof. Otherwise $\bigcup_{g\in G} \mathrm{Cen}(ij)^g \subseteq J^{\cdot 2}$ by Lemma 6.5, and hence $J^{\cdot 2}$ would be generic in G. This is impossible, since $\bigcup_{i\in J} \mathrm{Cen}(i) = \bigcup_{g\in G} H^g$ is generic in G and the elements of $J^{\cdot 2}$ do not have fixed points by Lemma 2.5. Therefore $\mathrm{Cen}(ij)$ has infinite index in $N_G(\mathrm{Cen}(ij))$.

Now $N_G(\mathrm{Cen}(ij)) = N_G(\ell_{ij})$, and $N_G(\ell_{ij}) \cap J = \ell_{ij}$ forms a mock hyperbolic reflection space (consisting of one line). Therefore $N_G(\mathrm{Cen}(ij)) = i\ell_{ij} N_{\mathrm{Cen}(i)}(\ell_{ij})$, and thus $N_{\mathrm{Cen}(i)}(\ell_{ij})$ is infinite. □

If the point stabilizer in a sharply 2-transitive group of characteristic $\neq 2$ with planar maximal near-field contains an element $g \notin \{1, i\}$ such that g normalizes all lines containing i, then by [Sozutov et al. 2014] the sharply 2-transitive group splits. We are going to prove a similar result for Frobenius groups of finite Morley rank of odd type.

If A is a group, then we write $A^* = A \setminus \{1\}$.

Lemma 6.17. *Let λ be a line containing $i \in J$ and fix a definable solvable subgroup $A \leq N_{\mathrm{Cen}(i)}(\lambda)$. Then $A^* i\lambda \cup \{1\} = A^\lambda$.*

Proof. Note that $H = \lambda^{\cdot 2} \rtimes A$ is a solvable Frobenius group of finite Morley rank. By [Borovik and Nesin 1994, Theorem 11.32], we have $H = \lambda^{\cdot 2} \cup \bigcup_{j\in\lambda} A^j$. This proves the lemma. □

Proposition 6.18. *Let Λ be a set of lines on J such that (J, Λ) forms a generic mock hyperbolic reflection space. Suppose there exists a definable infinite solvable normal subgroup $A \trianglelefteq \mathrm{Cen}(i)$ such that $A \leq \bigcap_{\lambda\in\Lambda_i} N_{\mathrm{Cen}(i)}(\lambda)$. Then $H < G$ splits.*

Proof. We may assume that all lines in Λ have Morley rank k. Since i is central in $\mathrm{Cen}(i)$, we may also assume that $i \in A$.

Now set $J_i = \bigcup_{\lambda\in\Lambda_i} \lambda = \{j \in J \setminus \{i\} : \ell_{ij} \text{ exists}\} \cup \{i\}$. By the previous lemma, we have $A^* i\lambda \cup \{1\} = A^\lambda$ for all $\lambda \in \Lambda_i$. Hence we have

$$A^* i J_i \cup \{1\} = \bigcup_{\lambda\in\Lambda_i} A^* i\lambda \cup \{1\} = \bigcup_{\lambda\in\Lambda_i} A^\lambda = A^{J_i}.$$

We have $A^* i J_i \approx A^* i J$ as a consequence of Lemma 2.5 and $A^{J_i} \approx A^J$, since J acts regularly on the set of conjugates of H and hence also on the set of conjugates of A.

Therefore

$$A^* i J \cup \{1\} \approx A^* i J_i \cup \{1\} = A^{J_i} \approx A^J = A^{\mathrm{Cen}(i)J} = A^G.$$

Put $N = \mathrm{Stab}^{\approx}(A^G)$. Then $A \leq N$ and $N \trianglelefteq G$ is a normal subgroup. Hence $A^G \leq N$. Now Proposition 5.4 implies that $A^G \approx N$. Note that

$$\mathrm{MR}(N) = \mathrm{MR}(J) + \mathrm{MR}(A).$$

Moreover, $J \subseteq N$, therefore $J^{\cdot 2} \subseteq N$ and thus $\mathrm{MR}(N) \geq 2n - k$. Note that A acts without fixed points on any line $\lambda \in \Lambda_i$, and therefore $\mathrm{MR}(A) \leq k$. In conclusion

$$n - k \leq \mathrm{MR}(N) - \mathrm{MR}(J) = \mathrm{MR}(A) \leq k,$$

and therefore $n \leq 2k$. Now Proposition 5.19 implies that $H < G$ splits. □

Corollary 6.19. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type. If H is a minimal group, i.e., if H does not contain an infinite proper definable subgroup, then $H < G$ splits.*

Proof. The assumptions and Lemma 6.16 imply that $N_{\mathrm{Cen}(i)}(\lambda) = \mathrm{Cen}(i)$ holds for all i in J and $\lambda \in \Lambda_i$. If $H = \mathrm{Cen}(i)$, then $H = \bigcap_{\lambda \in \Lambda_i} N_H(\lambda)$ and H is abelian. Therefore Proposition 6.18 implies that $H < G$ splits. $\square$

We can use Zilber's field theorem to find interpretable fields in Frobenius groups of odd type.

Proposition 6.20 [Borovik and Nesin 1994, Theorem 9.1]. *Let $G = A \rtimes H$ be a group of finite Morley rank, where A and H are infinite definable abelian subgroups and A is H-minimal, i.e., there are no definable infinite H-invariant subgroups. Assume that H acts faithfully on A. Then there is an interpretable field K such that $A \cong K_+$, $H \leq K^*$, and H acts by multiplication.*

Let $\lambda = \ell_{ij}$ be a line. Then $N_{\mathrm{Cen}(i)}(\lambda)$ is infinite and acts on $\lambda^{\cdot 2} = \mathrm{Cen}(ij)$ by conjugation. Take a minimal subgroup $A \leq N_{\mathrm{Cen}(i)}(\lambda)$. Since the action of A on $\mathrm{Cen}(ij)$ has no fixed points, we can find an infinite A-minimal subgroup $B \leq \mathrm{Cen}(ij)$ on which A acts faithfully. Moreover, B must be abelian, because $\mathrm{Cen}(ij)$ is an abelian group. Hence, by Proposition 6.20, there is an interpretable field K such that $B \cong K_+$, $A \leq K^*$, and A acts by multiplication.

In particular, if the line λ is strongly minimal, then K is strongly minimal and $A \cong K^*$.

If A is a proper subgroup of K^*, then K is a bad field, i.e., an infinite field of finite Morley rank such that K^* has a proper infinite definable subgroup. By [Baudisch et al. 2009], bad fields of characteristic 0 exist. However, it follows from work of Wagner [2001] that if $\mathrm{char}(K) \neq 0$, then K^* is a good torus, i.e., every definable subgroup of K^* is the definable hull of its torsion subgroup. We refer to [Cherlin 2005] for properties of these good tori.

Theorem 6.21. *Let $H < G$ be a connected Frobenius group of finite Morley rank and odd type. Fix Λ such that (J, Λ) is a generic mock hyperbolic reflection space. Moreover, assume that H has a definable nilpotent normal subgroup N such that $N \cap N_H(\lambda)$ is infinite for all $\lambda \in \Lambda_i$.*

If all lines in Λ are strongly minimal or if G does not interpret a bad field of characteristic 0, *then $H < G$ splits.*

Proof. We may assume that N is connected. Let T be a maximal good torus in N. As a consequence of the structure of nilpotent groups of finite Morley rank [Borovik and Nesin 1994, Theorems 6.8 and 6.9], T must be central in N. By [Cherlin 2005, Theorem 1], any two maximal good tori are conjugate. Therefore T is the unique maximal good torus in N. Since a connected subgroup of a good torus is a good torus, the assumptions (and the previous discussion) imply that $N_H(\lambda) \cap T$ is infinite for all lines $\lambda \in \Lambda_i$. By [Cherlin 2005, Lemma 2], the family $\{N_H(\lambda) \cap T : \lambda \in \Lambda_i\}$ is finite. Hence, after replacing Λ by a generic subset $\Lambda' \subseteq \Lambda$, we may assume that $\{N_H(\lambda) \cap T : \lambda \in \Lambda_i\}$ consists of a unique infinite abelian normal subgroup of H. Now Proposition 6.18 implies that $H < G$ splits. $\square$

Frobenius groups of degenerate type. We now use mock hyperbolic spaces to study Frobenius groups of finite Morley rank and degenerate type. A geometry with similar properties, but defined on the whole group, was used by Frécon in his result on the nonexistence of bad groups of Morley rank 3.

Lemma 6.22. *Let $H < G$ be a connected Frobenius group of Morley rank n and of degenerate type. Suppose the Frobenius complement H is abelian and of Morley rank k. Then $n \geq 2k+1$, and if $n = 2k+1$, then G contains a definable normal subgroup N of Morley rank $k+1$.*

Proof. Note that G is uniquely 2-divisible, and hence $a \otimes b = a^{1/2}ba^{1/2}$ defines a K-loop structure on G. Let $L = (G, \otimes)$ denote the corresponding K-loop, and set $\mathcal{A} = G \times \langle \epsilon \rangle < \mathrm{Aut}(L)$, where ϵ is given by inversion. Now let $\mathcal{G}$ be the quasidirect product $\mathcal{G} = L \rtimes_Q \mathcal{A}$.

By Theorem 4.7, the involutions J in $\mathcal{G}$ form a partial mock hyperbolic reflection space, and since $\bigcup_{g \in G} H^g \subseteq G$ is a generic subset of G, the involutions must form a generic mock hyperbolic reflection space. Moreover, $\mathrm{MR}(J) = n$ and each line has Morley rank k. Now the lemma follows from Theorem 5.5. $\square$

Theorem 6.23. *Let $H < G$ be a connected Frobenius group of Morley rank n and of degenerate type. Suppose the Frobenius complement H is abelian and of Morley rank k. Then $n \geq 2k+1$.*

If $n = 2k+1$, then G splits as $G = N \rtimes H$ for some definable connected normal subgroup N of Morley rank $k+1$. Moreover, if N is solvable, then there is an interpretable field K of characteristic $\neq 2$ such that $G = K_+ \rtimes H$, $H \leq K^*$, and H acts on K_+ by multiplication.

Proof. By the previous lemma, we may assume $n = 2k+1$. Then G contains a definable normal subgroup N of rank $k+1$, and we may assume that N is connected.

Note that $\mathrm{MR}\big(\bigcup_{g \in G}(N \cap H)^g\big) = k + 1 + \mathrm{MR}(N \cap H)$ and $\mathrm{MR}(N) = k+1$. Therefore $N \cap H$ must be finite. If $N \cap H$ is nontrivial, then $(N \cap H) < N$ is a connected Frobenius group, and hence $N \cap H$ must be connected. Therefore $N \cap H = \{1\}$.

The semidirect product $N \rtimes H$ has rank $2k+1$, and hence is generic in G. Therefore $G = N \rtimes H$ splits.

Now assume that N is solvable. Then N is nilpotent since, by [Borovik and Nesin 1994, Theorem 11.29], a solvable complement of a split Frobenius group of finite Morley rank is nilpotent. Moreover, $\mathrm{Cen}(u) \leq N$ for all $u \in N \setminus \{1\}$ by [Borovik and Nesin 1994, Theorem 11.32] (since G is solvable). Note that u^G cannot be generic in N, because G does not contain involutions. Therefore $\mathrm{MR}(u^G) \leq k$, and hence $\mathrm{MR}(\mathrm{Cen}(u)) \geq k+1$. Thus $\mathrm{Cen}(u) = N$, so N is abelian.

We now show that N is H-minimal: Let $A \leq N$ be a H-invariant subgroup. We may assume that A is connected. Given $a \in A \setminus \{1\}$, we have $\mathrm{Cen}(a) \cap H = \{1\}$, and therefore $a^H \subseteq A$ has rank k. If A has rank k, then a^H is generic in A, and therefore A must contain an involution. This is a contradiction. Therefore $A = \{1\}$ or $A = N$, and hence N is H-minimal.

By Proposition 6.20, there must be an interpretable field K such that $N = K_+$, $H \leq K^*$, and H acts on N by multiplication. □

7. Sharply 2-transitive groups of finite Morley rank

Let G be a sharply 2-transitive group of finite Morley rank with $\mathrm{char}(G) \neq 2$, and let J denote the set of involutions in G. By Corollary 3.4 (or by Theorem 6.11), the set J forms a mock hyperbolic reflection space.

We set $n = \mathrm{MR}(J)$ and $k = \mathrm{MR}(\mathrm{Cen}(ij))$ for involutions $i \neq j \in J$. Note that k does not depend on the choice of i, and j and $k = n$ if and only if G is split.

Now we assume that G is not split. By [Borovik and Nesin 1994, Proposition 11.71], we have $0 < 2k < n$, and we will improve this inequality below.

Since G acts sharply 2-transitively on J, it is easy to see that $\mathrm{MR}(G) = 2n$ and $\mathrm{MR}(J^{\cdot 2}) = 2n - k$. Moreover, G and $\mathrm{Cen}(ij)$ have Morley degree 1 by [Borovik and Nesin 1994, Lemma 11.60].

Proposition 7.1. (a) *The set iJ is indecomposable for all $i \in J$.*

(b) *$\langle J^{\cdot 2}\rangle$ is a definable connected subgroup. In particular, there is a bound m such that any $g \in \langle J^{\cdot 2}\rangle$ is a product of at most m translations.*

Proof. (a) Since $\mathrm{MD}(G) = 1$, the set J is indecomposable by [Borovik and Nesin 1994, Corollary 5.25], and hence iJ is indecomposable too.

(b) Since $\langle J^{\cdot 2}\rangle = \langle iJ\rangle$, (b) follows from Zilber's indecomposability theorem using (a). □

Remark 7.2. By Proposition 7.1(b), it is easy to see that the nonsplit examples of sharply 2-transitive groups of characteristic 0 constructed in [Rips and Tent 2019] do not have finite Morley rank.

Lemma 7.3. *For any $g \in G \setminus J$, the set $\{i \in J : gi \text{ has a fixed point}\}$ is generic in J.*

Proof. Let $g \in G$. For any $j \in J$ there is a unique $i_j \in J$ swapping j and j^g. Then gi_j centralizes j, so has a fixed point. If $i_j = i_p$ for some $j \neq p \in J$, then by sharp 2-transitivity it follows that $g = i_j = i_p \in J$. Hence for $g \notin J$, the i_j for $j \in J$ are pairwise distinct, and hence $\{i_j : j \in J\}$ has Morley rank n. □

Let $\mu : G^3 \to G$ be the multiplication map, i.e., $\mu(g_1, g_2, g_3) = g_1 g_2 g_3$.

Lemma 7.4. $$\mathrm{MR}(J^{\cdot 3}) > \mathrm{MR}(J^{\cdot 2}).$$

Proof. Note that $\mathrm{MR}(J^{\cdot 3}) = \mathrm{MR}(iJ^{\cdot 3}) \geq \mathrm{MR}(J^{\cdot 2}) > \mathrm{MR}(J) = n$, and hence J is not a generic subset of $J^{\cdot 3}$.

For $\alpha \in J^{\cdot 3}$, we let $X_\alpha = \{i \in J : i\alpha \in J^{\cdot 2}\}$ be the set of all involutions i such that $i\alpha$ is a translation. By Lemma 7.3 and Remark 3.1, $\mathrm{MR}(X_\alpha) < n$ for all $\alpha \in J^{\cdot 3} \setminus J$.

Let $\mathrm{MR}(J^{\cdot 3}) = 2n - k + l$ for some $l \geq 0$. There is a generic set of $\alpha \in J^{\cdot 3} \setminus J$ such that $\mathrm{MR}(\mu^{-1}(\alpha) \cap (J \times J \times J)) = n + k - l$. Set $X = X_\alpha$ for such an $\alpha \in J^{\cdot 3} \setminus J$. If $irs = \alpha$, then $\mathrm{MR}(\{j \in J : rs \in jJ\}) = k$, and hence $\mathrm{MR}(\mu^{-1}(\alpha)) = \mathrm{MR}(X) + k$. Therefore we have $\mathrm{MR}(X) = n - l$, and hence $l \geq 1$ by Lemma 7.3. □

Corollary 7.5. *Let G be a nonsplit sharply 2-transitive group of finite Morley rank. If the lines are strongly minimal, then G is simple and a counterexample to the Cherlin–Zilber conjecture.*

Proof. Let $N \trianglelefteq G$ be a normal subgroup. If N contains an involution, then $J \subseteq N$, and hence $J^{\cdot 2} \subseteq N$. Now assume N does not contain an involution. Fix $u \in N$ and $i \in J$. Then $1 \neq u^{-1}iui \in N \cap J^{\cdot 2}$, and hence $J^{\cdot 2} \subseteq N$, since all translations are conjugate. Therefore $J^{\cdot 2} \subseteq N$ holds true in both cases. Since $iJ^{\cdot 3} \subseteq \langle iJ \rangle \subseteq N$ and $\mathrm{MR}(J^{\cdot 2}) = 2n - 1 < \mathrm{MR}(J^{\cdot 3}) = \mathrm{MR}(iJ^{\cdot 3}) \leq \mathrm{MR}(G) = 2n$ (Lemma 7.4), this implies $N = G$. This shows that G must be simple.

Assume towards a contradiction that G is an algebraic group over an algebraically closed field K. If the K-rank of G is at least 2, then the torus contains commuting involutions, contradicting Remark 3.1(c). If the K-rank of G is 1, then G is isomorphic to $\mathrm{PSL}_2(K)$ and also contains commuting involutions, e.g., $x \mapsto -1/x$ and $x \mapsto -x$ are commuting involutions in $\mathrm{PSL}_2(K)$. □

Note that a sharply 2-transitive group of finite Morley rank in characteristic different from 2 is not a bad group in the sense of Cherlin, since for any translation $\sigma \in J^{\cdot 2}$ the group $N_G(\mathrm{Cen}(\sigma)) = \mathrm{Cen}(\sigma) \rtimes N_{\mathrm{Cen}(\sigma)}(\mathrm{Cen}(\sigma))$ is solvable, but not nilpotent.

If G is a sharply 2-transitive group of finite Morley rank and $\mathrm{char}(G) \neq 2$ with n, k and J be as before, then by Theorem 6.11, G splits if $n \leq 2k + 1$. Thus, we obtain:

Corollary 7.6. *If G is a sharply 2-transitive group and $\mathrm{MR}(G) = 6$, then G is of the form $\mathrm{AGL}_1(K)$ for some algebraically closed field K of Morley rank 3.*

Proof. If $\mathrm{char}(G) \neq 2$, then, by Theorem 6.11, G splits and the result follows from [Altınel et al. 2019]. If $\mathrm{char}(G) = 2$, then G is split by [Altınel et al. 2019] and any point stabilizer has Morley rank 3. Since the point stabilizers do not contain involutions, they are solvable by [Frécon 2018]. By [Borovik and Nesin 1994, Corollary 11.66], an infinite split sharply 2-transitive group of finite Morley rank whose point stabilizer contains an infinite normal solvable subgroup must be standard. □

8. Further remarks

A finite uniquely 2-divisible K-loop is the same as a finite B-loop in the sense of Glauberman [1964]. As a consequence of Glauberman's Z^*-theorem [1966] finite B-loops are solvable. Following Glauberman, we say that a K-loop L is *half-embedded* in some group G if it is isomorphic to a K-loop arising from a uniquely 2-divisible twisted subgroup of G as in Proposition 4.4. B-loops and uniquely 2-divisible K-loops can always be half-embedded in some group and that group can be chosen to be finite if the loop is finite [Glauberman 1964, Theorem 1 and Corollary 1]. This allows us to restate Glauberman's result for twisted subgroups:

Proposition 8.1 [Glauberman 1966]. *Let G be a group, and let $L \subseteq G$ be a finite uniquely 2-divisible twisted subgroup. Then $\langle L\rangle$ is solvable.*

As a consequence finite mock hyperbolic spaces must consist of a single line:

Proposition 8.2. *Suppose J forms a finite mock hyperbolic reflection space in a group G. Then J consists of a single line.*

Proof. We may assume that G acts faithfully on J. Let $i \in J$ be an involution. Since J acts regularly on itself, the square map on iJ must be injective and hence bijective as a consequence of finiteness. Now it is easy to check that iJ is a finite uniquely 2-divisible twisted subgroup in G. Therefore $\langle iJ\rangle$ is solvable by Proposition 8.1. Moreover, $\mathrm{Cen}(i) \leq N_G(\langle iJ\rangle)$ and G can be decomposed as $G = iJ\,\mathrm{Cen}(i)$. Therefore $\langle iJ\rangle$ is a solvable normal subgroup of G. It follows that G contains a nontrivial abelian normal subgroup. Now Theorem 2.14 implies that J consists of a single line. □

In the context of groups of finite Morley rank, we do not know if every uniquely 2-divisible K-loop of finite Morley rank can be definably half-embedded into a group of finite Morley rank. The following would be a finite Morley rank version of Glauberman's theorem:

Conjecture 8.3. Let G be a connected group of finite Morley rank with a definable uniquely 2-divisible twisted subgroup L of Morley degree 1 such that $G = \langle L\rangle$. Then G is solvable.

Note that this conjecture would imply the Feit–Thompson theorem for connected groups of finite Morley rank: if G is a connected group of finite Morley rank of degenerate type, then G is uniquely 2-divisible, and hence Conjecture 8.3 (applied to $L = G$) would imply that G is solvable.

Moreover, it would imply that Frobenius groups of finite Morley rank split: for Frobenius groups of degenerate type this would follow from solvability. If G is a connected Frobenius group of finite Morley rank and odd type with involutions J and lines Λ, then it suffices to show that G has a nontrivial definable solvable normal subgroup (in that case G has a nontrivial abelian normal subgroup and hence splits by Theorem 2.14). Note that iJ is a uniquely 2-divisible twisted subgroup. If G is sharply 2-transitive, then Proposition 7.1 shows that $\langle iJ\rangle$ is definable and connected and hence should be solvable by Conjecture 8.3.

For the general case consider the family $\mathcal{F}_i = \{\mathrm{Cen}(ij)^0 : j \in J \setminus \{i\}\}$. By Zilber's indecomposability theorem the subgroup $N = \langle H : H \in \mathcal{F}_i\rangle$ is definable and connected. Moreover, it is easy to see that $N \cap iJ$ must be generic in iJ and N must be normalized by $\mathrm{Cen}(i)$. Therefore N must be a normal subgroup of $G = iJ\,\mathrm{Cen}(i)$, and clearly $N = \langle N \cap iJ\rangle$. Therefore Conjecture 8.3 would imply that N is solvable.

If Frobenius groups of odd and degenerate type split, then Remark 6.3 shows that Frobenius groups of even type also split.

If the twisted subgroup in the statement of Conjecture 8.3 is strongly minimal, then we show that G must be 2-nilpotent:

Proposition 8.4. *Let G be a connected group of finite Morley rank with a definable strongly minimal uniquely 2-divisible twisted subgroup L such that $G = \langle L\rangle$. Then G is 2-nilpotent.*

Proof. Let $x \otimes y = x^{1/2}yx^{1/2}$ be the corresponding K-loop structure on L. If $(L, \otimes)$ is an abelian group, then [Kiechle 2002, Theorem 6.14, part (3)] implies $[[a, b], c] = 1$ for all $a, b, c \in L$, and therefore $G = \langle L\rangle$ must be 2-nilpotent. Therefore it suffices to show that $(L, \otimes)$ is an abelian group.

Put $T = N_G(L)/\mathrm{Cen}(L)$. Then $T \leq \mathrm{Aut}((L, \otimes))$, and we may consider the quasidirect product $\mathcal{G} = L \rtimes_Q T$. As stated in Proposition 4.5, the group $\mathcal{G} = L \rtimes_Q T$ acts transitively and faithfully on L by

$$(a, \alpha)(x) = a \otimes \alpha(x),$$

and T is the stabilizer of $1 \in L$. Note that $L' = L \times \{1\}$ is a uniquely 2-divisible twisted subgroup of $\mathcal{G}$. Hence $a \otimes' b = a^{1/2}ba^{1/2}$ defines a K-loop structure on L'. By [Kiechle 2002, Theorem 6.15], the K-loops $(L, \otimes)$ and $(L', \otimes')$ are isomorphic. Therefore it suffices to show that $(L', \otimes')$ is an abelian group.

Hrushovski's analysis of groups acting on strongly minimal sets [Borovik and Nesin 1994, Theorem 11.98] shows that $\mathrm{MR}(\mathcal{G}) \leq 3$. Moreover, if $\mathrm{MR}(\mathcal{G}) = 3$, then T acts sharply 2-transitively on $L \setminus \{1\}$, which is impossible, since T is a group of automorphisms of $(L, \otimes)$.

If $\mathrm{MR}(\mathcal{G}) = 2$, then $L \rtimes_Q T$ is a standard sharply 2-transitive group $K_+ \rtimes K^*$ (and the corresponding permutation groups coincide). Since L' acts without fixed points and the fixed-point-free elements of $K_+ \rtimes K^*$ are precisely the elements of K_+, L' is contained in K_+. Therefore $\otimes'$ agrees with the group structure on K_+, and hence $(L', \otimes')$ is an abelian group.

Now assume $\mathrm{MR}(\mathcal{G}) = 1$. We argue similarly to the proof of [Glauberman 1964, Lemma 5, part (v)].

Consider the finite twisted subgroup $L'' = \{a\mathcal{G}^0 : a \in L'\}$ of $\mathcal{G}/\mathcal{G}^0$. Since L' is uniquely 2-divisible, the map $L'' \to L''$, $a \mapsto a^2$ is surjective and hence a bijection, since L'' is finite. Hence we may define a K-loop structure $x\mathcal{G}^0 \otimes'' y\mathcal{G}^0 = x^{1/2}yx^{1/2}\mathcal{G}^0$ on L''. The natural map $L' \to L''$ is a surjective homomorphism from $(L', \otimes')$ to $(L'', \otimes'')$ with kernel $L' \cap \mathcal{G}^0$.

In particular, $L' \cap \mathcal{G}^0$ is a normal subloop of L'. Since $L'/(L' \cap \mathcal{G}^0)$ is finite and $\mathrm{MD}(L') = 1$, this implies $L' = L' \cap \mathcal{G}^0$, and hence $L' \subseteq \mathcal{G}^0$. The group $\mathcal{G}^0$ is strongly minimal and thus abelian. Therefore $\otimes'$ agrees with the group structure on $\mathcal{G}^0$, and therefore $(L', \otimes')$ is an abelian group. □

The proof of Proposition 8.4 in fact shows the following:

Corollary 8.5. *Let G be a group of finite Morley rank, and let $L \subseteq G$ be a definable uniquely 2-divisible twisted subgroup of G.*

(a) *If* $\mathrm{MD}(L) = 1$, *then* $L \subseteq G^0$.

(b) *If L is strongly minimal, then the associated K-loop $(L, \otimes)$ is an abelian group, and hence $\langle L\rangle$ is 2-nilpotent (without assuming that $\langle L\rangle$ is definable).*

In particular, if $(L, \otimes)$ is a strongly minimal uniquely 2-divisible K-loop such that L can be definably half-embedded into a group of finite Morley rank, then $(L, \otimes)$ is an abelian group.

Question 8.6. This suggests the following two questions:

(a) Suppose G and L satisfy the assumptions of Proposition 8.4. Must G be abelian?

(b) Is every strongly minimal (uniquely 2-divisible) K-loop an abelian group?

References

[Altınel et al. 2008] T. Altınel, 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

[Altınel et al. 2019] T. Altınel, A. Berkman, and F. O. Wagner, “Sharply 2-transitive groups of finite Morley rank”, preprint, 2019, available at https://hal.archives-ouvertes.fr/hal-01935537.

[Bachmann 1959] F. Bachmann, *Aufbau der Geometrie aus dem Spiegelungsbegriff*, Springer, 1959. 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

[Borovik and Nesin 1994] A. Borovik and A. Nesin, *Groups of finite Morley rank*, Oxford Logic Guides **26**, Clarendon, 1994. MR Zbl

[Borovik et al. 2007] 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

[Cherlin 2005] G. Cherlin, “Good tori in groups of finite Morley rank”, *J. Group Theory* **8**:5 (2005), 613–621. MR Zbl

[Frécon 2018] O. Frécon, “Simple groups of Morley rank 3 are algebraic”, *J. Amer. Math. Soc.* **31**:3 (2018), 643–659. MR Zbl

[Glauberman 1964] G. Glauberman, “On loops of odd order”, *J. Algebra* **1** (1964), 374–396. MR Zbl

[Glauberman 1966] G. Glauberman, “Central elements in core-free groups”, *J. Algebra* **4** (1966), 403–420. MR Zbl

[Jaligot 2001] E. Jaligot, “Full Frobenius groups of finite Morley rank and the Feit–Thompson theorem”, *Bull. Symbolic Logic* **7**:3 (2001), 315–328. MR Zbl

[Kerby 1974] W. Kerby, *On infinite sharply multiply transitive groups*, Hamburger Mathematische Einzelschriften **6**, Vandenhoeck & Ruprecht, Göttingen, 1974. MR Zbl

[Kiechle 2002] H. Kiechle, *Theory of K-loops*, Lecture Notes in Mathematics **1778**, Springer, 2002. MR Zbl

[Neumann 1940] B. H. Neumann, “On the commutativity of addition”, *J. London Math Soc.* **15** (1940), 203–208. 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

[Rips and Tent 2019] E. Rips and K. Tent, “Sharply 2-transitive groups of characteristic 0”, *J. Reine Angew. Math.* **750** (2019), 227–238. 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.* **19**:10 (2017), 2895–2910. MR Zbl

[Sozutov et al. 2014] A. I. Sozutov, E. B. Durakov, and E. V. Bugaeva, “On some near-domains and sharply doubly transitive groups”, *Tr. Inst. Mat. Mekh.* **20**:2 (2014), 277–283. In Russian. MR

[Tent and Ziegler 2016] K. Tent and M. Ziegler, “Sharply 2-transitive groups”, *Adv. Geom.* **16**:1 (2016), 131–134. MR Zbl

[Wagner 2001] F. Wagner, “Fields of finite Morley rank”, *J. Symbolic Logic* **66**:2 (2001), 703–706. MR Zbl

[Wagner 2017] F. Wagner, “Bad groups”, preprint, 2017. arXiv 1703.01764

Received 12 Jan 2022. Revised 19 Jun 2022.

TIM CLAUSEN:

tim.clausen@mailbox.org
Duesseldorf, Germany

KATRIN TENT:

tent@wwu.de
Universitaet Muenster, Muenster, Germany

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.177

Rigid differentially closed fields

David Marker

Using ideas from geometric stability theory we construct differentially closed fields of characteristic 0 with no nontrivial automorphisms.

1. Introduction

Our goal is to construct countable differentially closed fields of characteristic 0 (DCF_0) with no nontrivial automorphisms. We refer to such fields as *rigid*. This answers a question posed by Russell Miller. I will say something about Miller's motivation in my closing remarks. This may at first seem surprising. One often, naively, thinks that differentially closed fields should behave like algebraically closed fields, where there are always many automorphisms. Also, differential closures of proper differential subfields always have nontrivial automorphisms. We sketch the proof of this using ideas from Shelah's proof [18] of the uniqueness of prime models for ω-stable theories (see [12, §6.4] or [21, §9.2]). This is a well-known construction.

Proposition 1.1. *Let k be a differential field with differential closure $K \supset k$. Then there are nontrivial automorphisms of K/k.*

Proof. First note that if $\boldsymbol{d} \in K^n$ and $k\langle \boldsymbol{d}\rangle$ is the differential field generated by $\boldsymbol{d}$ over k, then K is a differential closure of $k\langle \boldsymbol{d}\rangle$. This follows from the fact that in an ω-stable theory $\mathcal{M}$ is prime over $A \subset \mathcal{M}$ if and only if $\mathcal{M}$ is atomic over A and there are no uncountable sets of indiscernibles (see [21, Theorem 9.2.1]).

Let $a \in K \setminus k$. Since K is the differential closure of k, $\mathrm{tp}(a/k)$ is isolated by some formula $\phi(v)$ with parameters from k. If a is the only element of K satisfying ϕ, then a is in $\mathrm{dcl}(k) = k$, a contradiction. Thus there is $b \in K$ such that $a \neq b$ and $\phi(b)$.

Since a and b realize the same type over k, there are $L \models \mathrm{DCF}_0$ with $k\langle b\rangle \subseteq L$ and $\sigma : K \to L$ an isomorphism such that $\sigma \mid k$ is the identity and $\sigma(a) = b$.

The author was partially supported by the Fields Institute for Research in the Mathematical Sciences.
MSC2020: 03C45, 03C60, 12H05.
Keywords: model theory, differential fields, geometric stability theory.

K is a differential closure of both $k\langle a\rangle$ and $k\langle b\rangle$. Thus L is a differential closure of $k\langle b\rangle$ and, by uniqueness of differential closures, there is an isomorphism $\tau : L \to K$ that is the identity on $k\langle b\rangle$. Then $\tau \circ \sigma$ is an automorphism of K sending a to b. □

Remarks. • This argument really shows that if T is an ω-stable theory, A is a definably closed substructure of a model of T that is not a model of T and $\mathcal{M}$ is a prime model extension of A, then there is a nontrivial automorphism of $\mathcal{M}$ fixing A pointwise.

• While this argument guarantees the existence of a nontrivial automorphism of K/k, it is possible that there is only one. If k is a model of Singer's theory of *closed ordered differential fields* [20], then $k^{\text{diff}} = k(i)$ and complex conjugation is the only nontrivial automorphism of k^{diff}/k.

Omar León Sánchez pointed out that the construction of rigid differentially closed fields gives the first known examples of differentially closed fields K such that $K \neq k(i)$ for any closed ordered differential field $k \subset K$.

• Proposition 1.1 tells us that the rigid differentially closed fields we construct are not the differential closure of any proper differential subfield.

Our construction of rigid differentially closed fields uses ideas from geometric stability theory and work on strongly minimal sets in differentially closed fields of Rosenlicht [17] and Hrushovski and Sokolović [9]. We describe the results we need in Section 2 and construct rigid differentially closed fields in Section 3. We begin Section 3 with a warm up constructing arbitrarily large rigid models and then give the more subtle construction of rigid countable models. We refer the reader to [15] for unexplained model theoretic concepts.

2. Preliminaries

We work in $\mathbb{K} \models \text{DCF}$, a monster model of the theory of differentially closed fields of characteristic zero with a single derivation. The constant field C is $\{x \in \mathbb{K} : x' = 0\}$. If k is a differential field and $X \subset \mathbb{K}^n$ is definable over k, we let $X(k)$ denote the k-points of X, i.e., $X(k) = k^n \cap X$. Of course, by quantifier elimination, X is quantifier-free definable over k

Our main tool will be the strongly minimal sets known as *Manin kernels* of elliptic curves. Manin kernels arose in Manin's proof [10] of the Mordell conjecture for function fields in characteristic zero and were central to both Buium's [2] and Hrushovski's [8] proofs of the Mordell–Lang conjecture for function fields in characteristic zero. The model theoretic importance of Manin kernels was developed in the beautiful unpublished preprint of Hrushovski and Sokolović [9]. Proofs of the results from [9] that we will need all appear in Pillay's survey [16], and [11] is another survey on the construction and some of the basic properties of Manin kernels.

For $a \in K$, let E_a be the elliptic curve $Y^2 = X(X-1)(X-a)$. Let $E_a^\sharp$ be the minimal definable differential subgroup of E. $E_a^\sharp$ is the closure of $\mathrm{Tor}(E_a)$ in the Kolchin topology.

Theorem 2.1 (Hrushovski–Sokolović). (i) *If $a' \neq 0$, then $E_a^\sharp$ is a nontrivial locally modular strongly minimal set.*

(ii) *The Manin kernels $E_a^\sharp$ and $E_b^\sharp$ are nonorthogonal if and only if E_a and E_b are isogenous. In particular, if a and b are algebraically independent over $\mathbb{Q}$ then $E_a^\sharp$ and $E_b^\sharp$ are orthogonal.*

In particular, Manin kernels are orthogonal to the field of constants $C = \{x : x' = 0\}$.

More generally, if A is a simple abelian variety that is not isomorphic to an abelian variety defined over the constants we can construct a Manin kernel $A^\sharp$ which is the Kolchin closure of the torsion of A and a minimal infinite definable subgroup of A. $A^\sharp$ is nontrivial locally modular strongly minimal and Hrushovski and Sokolović also showed that if X is any nontrivial locally modular strongly minimal subset of a differentially closed field, then $X \not\perp A^\sharp$ for some abelian variety A.

The other building blocks of our construction are strongly minimal sets introduced by Rosenlicht [17] in his proof that the differential closure of a differential field k need not be minimal.

Let $f(X) = X/(1+X)$. For $a \neq 0$, let $X_a = \{x : x' = af(x), x \neq 0\}$.

Theorem 2.2 (Rosenlicht). (i) *If $a \in k$ and $x \in X_a \setminus k$, then $C(k) = C(k\langle x\rangle)$.*

(ii) *Suppose $k \subset K$ are differential fields, with $C(K) \subseteq C(k)^{\mathrm{alg}}$. Suppose $a, b \in k^\times$, $x \in X_a(K)$, $y \in X_b(K)$ and x and y are algebraically dependent over k. Then x, y are algebraic over k or $x = y$. In particular, if $a \neq b$, then X_a and X_b are orthogonal.*

Part (i) follows from Proposition 2 of [17] while (ii) is a slight generalization of Proposition 1 of [17] and Gramain [5]. These results appear as Theorems 6.12 and 6.2 of [13].

Corollary 2.3. *Each X_a is a trivial strongly minimal set.*

Proof. By Theorem 2.2(i), X_a is orthogonal to the constants. If X_a were nontrivial, then $X_a \not\perp A^\sharp$, the Manin kernel of a simple abelian variety. But if $x \in X_a \setminus k^{\mathrm{alg}}$, then $k\langle x\rangle = k(x)$ is a transcendence degree 1 extension. But by results of Buium [2], Manin kernels, or anything nonorthogonal to one, give rise to extensions of transcendence degree at least 2. Thus X_a is trivial. □

3. Constructing rigid differentially closed fields

Warm up.

Proposition 3.1. *There are arbitrarily large rigid differentially closed fields.*

For this construction we only need Rosenlicht strongly minimal sets. Let κ be a cardinal with $\kappa = \aleph_\kappa$. We construct a differentially closed field K of cardinality κ such that $|X_a(K)| \neq |X_b(K)|$ for each nonzero $a \neq b$, guaranteeing there is no automorphism sending $a \mapsto b$.

We build a chain of differentially closed fields $K_0 \subset K_1 \subset \cdots \subset K_\alpha \subset \cdots$ for $\alpha < \kappa$ such that $|K_\alpha| = \aleph_\alpha$. We simultaneously build an injective enumeration $a_0, a_1, \ldots, a_\alpha, \ldots$ of $K^\times$, where $K = \bigcup K_\alpha$.

We construct K as follows.

(i) $K_0 = \mathbb{Q}^{\text{diff}}$.

(ii) Given K_α and $a_\alpha \in K_\alpha$, build $K_{\alpha+1}$ by adding $\aleph_{\alpha+1}$ new independent elements of X_{a_α} and taking the differential closure.

(iii) If α is a limit ordinal, let $K_\alpha = \bigcup_{\beta<\alpha} K_\beta$.[1]

Since $X_{a_\alpha} \perp X_{a_\beta}$ for $\alpha < \beta$, adding new elements to X_{a_β} and taking the differential closure adds no new elements to X_{a_α}. Thus $X_{a_\alpha}(K) = X_{a_\alpha}(K_{\alpha+1})$. In particular, $|X_{a_\alpha}(K)| = \aleph_{\alpha+1}$. Thus there is no automorphism of K with $a_\alpha \mapsto a_\beta$ for $\alpha \neq \beta$.

One might worry that we have contradicted Proposition 1.1. Let B_α be all of the independent realizations of X_{a_α} that we added at stage α. Then K is the differential closure of $k = \mathbb{Q}\langle B_\alpha : \alpha < \kappa\rangle$. But, if $b \in X_{a_\alpha}$, then $a_\alpha = b'(b+1)/b \in \mathbb{Q}\langle b\rangle$. Thus $k = K$.

The countable case. To construct a countable differentially closed field with no automorphisms, we need a more subtle mixture of Rosenlicht extensions with extensions of Manin kernels.

Suppose $b \notin C$. Let $\dim E_b^\sharp(k)$ be the number of independent realizations in k of the generic type of $E_b^\sharp$ over $\mathbb{Q}\langle b\rangle$. Manin kernels are useful to us as they can have any countable dimension. We build a countable $K \models \mathrm{DCF}_0$ such that for each $a \neq 0$, there is a natural number

$$n_a = \max_{b \in X_a(K)} \dim E_b^\sharp(K)$$

such that $n_a \neq n_b$ for $a \neq b$. This guarantees that there is no automorphism with $a \mapsto b$.

[1] To build the desired enumeration, let $a_0, a_1, \ldots$ be an injective enumeration of K_0 and, at stage $\alpha + 1$, let $(a_\gamma : \omega_\alpha \leq \gamma < \omega_{\alpha+1})$ be an injective enumeration of $K_{\alpha+1} \setminus K_\alpha$.

Freitag and Scanlon [4], and more generally, Casale, Freitag and Nagloo [3], have given constructions of trivial strongly minimal sets which can take on any countable dimension. Presumably these could be used in an alternative construction.

We build a chain $K_0 \subset K_1 \subset \cdots \subset K_n \subset \cdots$, an injective enumeration $a_0, a_1, \ldots$ of $K^\times = \bigcup K_n^\times$ and a sequence of natural numbers $0 = n_0 < n_1 < \cdots$ such that

(1) $C(K_i) = C(K_0)$;

(2) $X_{a_i}(K) = X_{a_i}(K_{i+1})$;

(3) if $b \in X_{a_i}(K)$, then $E_b^\sharp(K) = E_b^\sharp(K_{i+1})$;

(4) $n_{i+1} = \max_{b \in X_{a_i}(K)} \dim E_b^\sharp$.[2]

If we can do that we will have guaranteed that there are no automorphisms of K.

Let $K_0 = \mathbb{Q}^{\text{diff}}$. At stage s we choose a new $a_s \in K_s$. Let b_s be an element of X_{a_s} generic over K_s, let $\boldsymbol{x}$ be $n_{s-1}+1$ independent realizations of the generic of $E_{b_s}^\sharp$ over $K_s\langle b_s\rangle = K_s(b_s)$ and let $K_{s+1} = K_s\langle b_s, \boldsymbol{x}\rangle^{\text{diff}}$.

By orthogonality considerations, it's clear that conditions (1)–(3) hold, as after stage $i+1$ we only add realizations of types orthogonal to X_{a_i} and $E_b^\sharp$, for $b \in X_{a_i}(K)$. To prove (4) we need to show that there is $n_s = \max_{d \in X_{a_s}} \dim E_d^\sharp(K_{s+1})$. We have arranged things so that if there is a bound n_s then $n_s > n_{s-1}$.

We need two preliminary lemmas.

Lemma 3.2. *If* $b' \neq 0$, *then* $\dim E_b^\sharp(\mathbb{Q}\langle b\rangle^{\text{diff}}) = 0$.

Proof. Suppose $x \in E_b^\sharp(\mathbb{Q}\langle b\rangle^{\text{diff}})$. All torsion points of E_b are in $\mathbb{Q}(b)^{\text{alg}}$, so we can suppose x is a nontorsion point. But x realizes an isolated type over $\mathbb{Q}\langle b\rangle$. Let ψ isolate the type of x over $\mathbb{Q}\langle b\rangle$. No torsion point can satisfy ψ. Thus by strong minimality ψ defines a finite set and $x \in \mathbb{Q}\langle b\rangle^{\text{alg}}$. □

Although we do not need it, we can say more in the special case that $\mathbb{Q}\langle b\rangle = \mathbb{Q}(b)$, such as if $b \in X_a$ for some $a \in \mathbb{Q}$. In this case Manin's theorem of the kernel [10] implies that $E_b^\sharp(\mathbb{Q}\langle b\rangle^{\text{alg}}) = \text{Tor}(E_b)$; see [1, Corollary K.3].

Lemma 3.3. *Suppose K is a differentially closed field, $b, d \in K$ and E_b and E_d are isogenous. Then* $\dim E_b^\sharp(K) = \dim E_d^\sharp(K)$.

Proof. If E_d and E_b are isogenous, then d and b are interalgebraic over $\mathbb{Q}$ and the isogeny f is defined over $\mathbb{Q}(d)^{\text{alg}} = \mathbb{Q}(b)^{\text{alg}}$. Since $f : \text{Tor}(E_d) \to \text{Tor}(E_b)$ is finite-to-one and the torsion is Kolchin dense in a Manin kernel, $f : E_d^\sharp \to E_b^\sharp$ is finite-to-one. It follows that $\dim E_d^\sharp(K) = \dim E_b^\sharp(K)$. □

The next lemma shows that we have the necessary bounds.

[2] Building the enumeration takes a bit more bookkeeping in this case. Let $d_{0,0}, d_{0,1}, \ldots$ be an injective enumeration of K_0 and let $d_{i,0}, d_{i,1}, \ldots$ be an injective enumeration of $K_i \setminus K_{i-1}$. Start our enumeration of K by letting $a_0 = d_{0,0}$. Suppose we start stage i with the partial enumeration $a_0, \ldots, a_M$. Then for $j = 0, \ldots, i$, let $a_{M+j+1} = d(i, i-j)$.

Lemma 3.4. *Suppose K is a differentially closed field constructed in a finite iteration $\mathbb{Q}^{\text{diff}} = k_0 \subset k_1 \subset \cdots \subset k_m = K$, where either*

(1) $k_{i+1} = k_i\langle a\rangle^{\text{diff}}$, *where a realizes a trivial type over k_i, or*

(2) $k_{i+1} = k_i\langle \boldsymbol{x}_i\rangle^{\text{diff}}$, *where $\boldsymbol{x}_i$ consists of n_i independent realizations of the generic type of a Manin kernel $E^\sharp_{b_i}$, where $b_i \in k_i$ and $E^\sharp_{b_i} \perp E^\sharp_{b_j}$ for $i \neq j$.*

If $d \in K \setminus C$, then $\dim E^\sharp_d(K) = n_i$ *for some i.*

Proof. We first argue that this is true for each $E^\sharp_{b_t}$. Define $l_0 \subseteq l_1 \subseteq \cdots \subseteq l_t$ such that $l_i = k_i\langle b_t\rangle^{\text{diff}}$. Note that $l_t = k_t$

By Lemma 3.2, $\dim E^\sharp_{b_t}(l_0) = 0$. As we construct $l_1, \ldots, l_t$ we are either doing nothing (if a_i or $\boldsymbol{x}_i \in l_{i-1}$) or adding realizations of types orthogonal to $E^\sharp_{b_t}$. Thus $\dim E^\sharp_{b_t}(k_t) = 0$ and $\dim E^\sharp_{b_t}(k_{t+1}) = n_t$. Since for $i > t$ all a_i and $\boldsymbol{x}_i$ realize types orthogonal to $E^\sharp_{b_t}$, $\dim E^\sharp_{b_t}(K) = n_t$.

Suppose $d \in K \setminus C$. If E_d is isogenous to some E_{b_i}, then, by Lemma 3.3, $\dim E^\sharp_d(K) = \dim E^\sharp_{b_i}(K) = n_i$. So we may assume $E^\sharp_d \perp E^\sharp_{b_i}$ for all i. We claim that in this case, $\dim E^\sharp_d(K) = 0$. For $i \leq m$, we let $l_i = k_i\langle d\rangle^{\text{diff}}$. By Lemma 3.2, $\dim E^\sharp_d(l_0) = 0$. As we continue the construction, as above, at each stage we either do nothing or realize types that are orthogonal to $E^\sharp_d$. Thus we add no new elements of $E^\sharp_d$ and $\dim E^\sharp_d(K) = 0$. □

We can interweave a many models construction. In [9] the authors noted that Manin kernels could be used to show that DCF_0 has eni-dop and concluded that there are $2^{\aleph_0}$ nonisomorphic countable differentially closed fields. An explicit version of this construction coding graphs into models is used in [14]. We can fold that coding into our construction of a rigid model.

Theorem 3.5. *There are $2^{\aleph_0}$ nonisomorphic countable rigid differentially closed fields. Each of these fields is not the differential closure of a proper differential subfield.*

Consider $X = X_1(\mathbb{Q}^{\text{diff}})$. This is an infinite set of algebraically independent elements. Let $G = (X, R)$ be a graph with vertex set X and edge relation R. Let $(\{u_i, v_i\} : i = 0, 1, \ldots)$ be an enumeration of two element subsets of X. We modify our construction such that at stage s we also add a generic element of $E^\sharp_{u_i+v_i}$ if and only if $(u_i, v_i) \in R$. We can still apply Lemma 3.4 and our construction will produce a rigid differentially closed field K. From K we can recover the graph in an $\mathcal{L}_{\omega_1,\omega}$-definable way. Thus nonisomorphic graphs give rise to nonisomorphic rigid differentially closed fields.

Similarly, we could interweave graph coding steps in the proof of Proposition 3.1 and build 2^κ nonisomorphic rigid differentially closed fields of cardinality κ when $\kappa = \aleph_\kappa$.

4. Remarks and Questions

In [6; 7] the authors introduce the notion of computable and Borel functors between classes of countable structures. For example, in Theorem 3.5, recovering the graph from the differentially closed field is a Borel functor from differentially closed fields to graphs. Miller wondered if there could be invertible functors between these classes. If there is an invertible functor F from graphs to differentially closed fields, then the authors show that the corresponding automorphism groups $\mathrm{Aut}(G)$ and $\mathrm{Aut}(F(G))$ would be isomorphic. Miller's original idea was that, since there are rigid graphs, one could show there was no such functor by showing that there are no rigid differentially closed fields. While our construction shows that this idea does not work, nevertheless, one can show there is no such functor by looking at possible automorphism groups. It is easy to construct a countable graph with an automorphism of order $n > 2$. But no differentially closed field can have an automorphism of order $n > 2$. Suppose K is differentially closed and σ is an automorphism of order $n > 2$. Let F be the fixed field of σ. Then K/F is an algebraic extension of order $n > 2$. By the Artin–Schreier theorem, this is impossible for K algebraically closed.

Question 1. Is there a differentially closed field K where $|\mathrm{Aut}(K)| = 2$? If so, is the fixed field a model of CODF? More generally, if K is a real closed differential field and $K(i)$ is differentially closed, must K be a model of CODF?

Question 2. Are there rigid differentially closed fields of cardinality $\aleph_1$?

The construction of such a model would require a new strategy. Perhaps it would help to assume the set theoretic principle $\diamondsuit$? Or perhaps one could use the methods of [19].

Acknowledgments

I am grateful to Russell Miller for bringing this question to my attention and to Zoé Chatzidakis, Jim Freitag, Omar León Sánchez and the referee for remarks on earlier drafts.

I am pleased to submit this paper in honor of Udi Hrushovski's belated 60th birthday. The main result relies heavily on his work.

References

[1] D. Bertrand and A. Pillay, "A Lindemann–Weierstrass theorem for semi-abelian varieties over function fields", *J. Amer. Math. Soc.* **23**:2 (2010), 491–533. MR Zbl

[2] A. Buium, *Differential algebra and Diophantine geometry*, Hermann, Paris, 1994. MR Zbl

[3] 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

[4] J. Freitag and T. Scanlon, "Strong minimality and the j-function", *J. Eur. Math. Soc.* (*JEMS*) **20**:1 (2018), 119–136. MR Zbl

[5] F. Gramain, "Non-minimalité de la clôture différentielle, II: La preuve de M. Rosenlicht", exposé 5 in *Groupe d'étude de théories stables* (*Bruno Poizat*), 3, 1980/1982, Univ. Paris VI, Paris, 1983. MR Zbl

[6] M. Harrison-Trainor, A. Melnikov, R. Miller, and A. Montalbán, "Computable functors and effective interpretability", *J. Symb. Log.* **82**:1 (2017), 77–97. MR Zbl

[7] M. Harrison-Trainor, R. Miller, and A. Montalbán, "Borel functors and infinitary interpretations", *J. Symb. Log.* **83**:4 (2018), 1434–1456. MR Zbl

[8] E. Hrushovski, "The Mordell–Lang conjecture for function fields", *J. Amer. Math. Soc.* **9**:3 (1996), 667–690. MR Zbl

[9] E. Hrushovski and Z. Sokolović, "Minimal subsets of differentially closed fields", preprint, 1994.

[10] Y. I. Manin, "Proof of an analogue of Mordell's conjecture for algebraic curves over function fields", *Dokl. Akad. Nauk SSSR* **152** (1963), 1061–1063. In Russian. MR

[11] D. Marker, "Manin kernels", pp. 1–21 in *Connections between model theory and algebraic and analytic geometry*, edited by A. Macintyre, Quad. Mat. **6**, Dept. Math., Seconda Univ. Napoli, Caserta, 2000. MR Zbl

[12] D. Marker, *Model theory*, Grad. Texts in Math. **217**, Springer, 2002. MR Zbl

[13] D. Marker, "Model theory of differential fields", pp. 41–113 in *Model theory of fields*, 2nd ed., Lect. Notes in Logic **5**, Association for Symbolic Logic, La Jolla, CA, 2006. MR Zbl

[14] D. Marker and R. Miller, "Turing degree spectra of differentially closed fields", *J. Symb. Log.* **82**:1 (2017), 1–25. MR Zbl

[15] A. Pillay, *Geometric stability theory*, Oxford Logic Guides **32**, Oxford Univ. Press, 1996. MR Zbl

[16] A. Pillay, "Differential algebraic groups and the number of countable differentially closed fields", pp. 114–134 in *Model theory of fields*, 2nd ed., Lect. Notes in Logic **5**, Association for Symbolic Logic, La Jolla, CA, 2006. MR Zbl

[17] M. Rosenlicht, "The nonminimality of the differential closure", *Pacific J. Math.* **52** (1974), 529–537. MR Zbl

[18] S. Shelah, "Differentially closed fields", *Israel J. Math.* **16** (1973), 314–328. MR Zbl

[19] S. Shelah, "Models with second order properties, IV: A general method and eliminating diamonds", *Ann. Pure Appl. Logic* **25**:2 (1983), 183–212. MR Zbl

[20] M. F. Singer, "The model theory of ordered differential fields", *J. Symbolic Logic* **43**:1 (1978), 82–91. MR Zbl

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

Received 17 Jan 2022. Revised 16 Dec 2022.

DAVID MARKER:

marker@uic.edu
Mathematics, Statistics, and Computer Science, University of Illinois Chicago, Chicago, IL, United States

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.185

Definable convolution and idempotent Keisler measures, II

Artem Chernikov and Kyle Gannon

With gratitude to Ehud Hrushovski, whose beautiful ideas have deeply influenced the authors.

We study convolution semigroups of invariant/finitely satisfiable Keisler measures in NIP groups. We show that the ideal (Ellis) subgroups are always trivial and describe minimal left ideals in the definably amenable case, demonstrating that they always form a Bauer simplex. Under some assumptions, we give an explicit construction of a minimal left ideal in the semigroup of measures from a minimal left ideal in the corresponding semigroup of types (this includes the case of $\mathrm{SL}_2(\mathbb{R})$, which is not definably amenable). We also show that the canonical pushforward map is a homomorphism from definable convolution on $\mathcal{G}$ to classical convolution on the compact group $\mathcal{G}/\mathcal{G}^{00}$, and use it to classify $\mathcal{G}^{00}$-invariant idempotent measures.

1. Introduction

This paper is a continuation of [Chernikov and Gannon 2022], but with a focus on NIP groups and the dynamical systems associated to the definable convolution operation. It was demonstrated in [Chernikov and Gannon 2022] that when T is an NIP theory expanding a group, $\mathcal{G}$ is a monster model of T, and $G \prec \mathcal{G}$, the spaces of global $\mathrm{Aut}(\mathcal{G}/G)$-invariant Keisler measures and Keisler measures which are finitely satisfiable in G (denoted by $\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ and $\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, G)$, respectively) form left-continuous compact Hausdorff semigroups under definable convolution $*$ (see Fact 2.29). Equivalently, the semigroup $(\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, G), *)$ can be described as the Ellis semigroup of the dynamical system given by the action of $\mathrm{conv}(G)$, the convex hull of G in the space of global measures finitely satisfiable in G, on the space of measures $\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, G)$ (see [Chernikov and Gannon 2022, Theorem 6.10 and Remark 6.11]). The main purpose of this paper is to study the structure of these semigroups, as well as to provide a description of idempotent measures via type-definable subgroups in some cases.

In Section 2 we review some preliminaries and basic facts on convolution in compact topological groups (Section 2A), model theory (Section 2B), Keisler

MSC2020: primary 03C45, 37B05, 43A10; secondary 03C60, 28D15.
Keywords: Keisler measures, convolution, idempotent measures, NIP, Ellis group, minimal ideals.

measures (Section 2C), definable convolution in NIP groups (Section 2D), Ellis semigroups (Section 2E) and Choquet theory (Section 2F).

In Section 3 we study the relationship between the semigroups $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ and $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ (under definable convolution) and the classical convolution semigroup of regular Borel probability measures on the compact topological group $\mathcal{G}/\mathcal{G}^{00}$. We demonstrate that the pushforward along the quotient map is a surjective, continuous, semigroup homomorphisms from definable convolution to classical convolution on $\mathcal{G}/\mathcal{G}^{00}$ (Theorem 3.10), mapping idempotent Keisler measures onto idempotent Borel measures on $\mathcal{G}/\mathcal{G}^{00}$ (Corollaries 3.11 and 3.12).

We have shown in [Chernikov and Gannon 2022, Theorem 5.8] that, by analogy to the classical theorem of Kawada and Itô for compact groups (Fact 2.8), which was later rediscovered by Wendel, there is a one-to-one correspondence between idempotent measures on a stable group and its type-definable subgroups (namely, every idempotent measure is the unique translation-invariant measure on its type-definable stabilizer group). In NIP groups, this correspondence fails (Example 4.5), but revised versions of this statement can be recovered in some cases. In particular, using the results of Section 3, we demonstrate in Section 4 that a $\mathcal{G}^{00}$-invariant idempotent measure in an NIP group $\mathcal{G}$ is a (not necessarily unique) invariant measure on its type-definable stabilizer group. In future work, we examine further cases of the classification of idempotent measures in NIP groups, including the generically stable case.

In Section 5 we study the semigroups $(\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G), *)$ and $(\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G), *)$ for an NIP group $\mathcal{G}$ through the lens of Ellis theory. We demonstrate that the ideal subgroups of any minimal left ideal (in either $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ or $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$) are always trivial (Proposition 5.10). This is due to the presence of the convex structure, in contrast to the case of types in definably amenable NIP groups (where, due to the proof of the Ellis group conjecture in [Chernikov and Simon 2018], the ideal subgroups are isomorphic to $\mathcal{G}/\mathcal{G}^{00}$). We also classify minimal left ideals in both $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ and $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ when $\mathcal{G}$ is definably amenable. In this case, any minimal left ideal in $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ is also trivial (Proposition 5.16), while $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ contains a unique minimal left ideal (which is also two-sided). This unique ideal is precisely the collection of measures in $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ which are $\mathcal{G}$-right-invariant (Proposition 5.18; this is in contrast to minimal left ideals in $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ corresponding to *G-left*-invariant measures). It is also a compact convex set, and moreover a *Bauer simplex* (see Definition 2.38). In particular, the set of its extreme points is closed, and consists precisely of the lifts μ_p of the Haar measure on $\mathcal{G}/\mathcal{G}^{00}$ for $p \in S_x^{\mathrm{inv}}(\mathcal{G}, G)$ an f-generic type of $\mathcal{G}$ (Corollary 5.21). If the group $\mathcal{G}$ is fsg, this minimal ideal is also trivial (Corollary 5.24). We also observe that if $\mathcal{G}$ is not definably amenable, then the minimal left ideal of $\mathfrak{M}_x(\mathcal{G}, G)$ has infinitely many extreme points (Remark 5.26). See Theorem 5.1 for a more precise summary of the results of the section.

In Section 6 we isolate certain conditions on $\mathcal{G}$, applicable in particular to some nondefinably amenable groups, which allows us to describe a minimal left ideal of $\mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$ for $\dagger \in \{\mathrm{fs}, \mathrm{inv}\}$ in terms of a minimal ideal in the corresponding semigroup of types. We prove the following two results. Suppose that I is a minimal left ideal of $S^{\dagger}_{x}(\mathcal{G}, G)$ and u is an idempotent in I such that $u * I$ is a compact group under the induced topology (we refer to this condition as CIG1; see Definition 6.5). Then $\mathfrak{M}(I) * \mu_{u*I}$ is a minimal left ideal of $\mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$, where μ_{u*I} is the Keisler measure corresponding to the normalized Haar measure on $u * I$ and $\mathfrak{M}(I)$ is the set of Keisler measures supported on I (Theorem 6.11). Under a stronger assumption, CIG2, on $\mathcal{G}$ (see Definition 6.14), we show that a minimal left ideal of $\mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$ is affinely homeomorphic to a collection of regular Borel probability measures over a natural quotient of I; specifically, it is a Bauer simplex (Theorem 6.20). In particular, $\mathrm{SL}_2(\mathbb{R})$ falls into both of these categories (Example 6.23).

2. Preliminaries

Given $r_1, r_2 \in \mathbb{R}$ and $\varepsilon \in \mathbb{R}_{>0}$, we write $r_1 \approx_{\varepsilon} r_2$ if $|r_1 - r_2| < \varepsilon$. For $n \in \mathbb{N}_{\geq 1}$, $[n] = \{1, 2, \ldots, n\}$.

2A. ***The classical setting.*** Before discussing the model-theoretic setting, we recall some classical facts concerning compact Hausdorff spaces, measures, and compact topological groups.

Fact 2.1. *Let X, Y be compact Hausdorff spaces and $f : X \to Y$.*

(i) *Let $\mathcal{M}(X)$ be the set of all regular Borel probability measures on X. Then $\mathcal{M}(X)$ is a compact Hausdorff space under the weak-* topology, with the basic open sets of the form*

$$\bigcap_{i=1}^{n} \left\{ \mu \in \mathcal{M}(X) : r_i < \int_X f_i \, d\mu < s_i \right\}$$

for $n \in \mathbb{N}$, $r_i < s_i \in \mathbb{R}$ and $f_i : X \to \mathbb{R}$ continuous for $i \in [n]$.

(ii) *A net of measures $(\mu_i)_{i \in I}$ in $\mathcal{M}(X)$ converges to a measure μ if and only if for any continuous $f : X \to \mathbb{R}$,*

$$\lim_{i \in I} \int_X f \, d\mu_i = \int_X f \, d\mu.$$

(iii) *A (Borel) measurable map $f : X \to Y$ induces the **pushforward** map*

$$f_* : \mathcal{M}(X) \to \mathcal{M}(Y)$$

given by $f_*(\mu)(A) = \mu(f^{-1}(A))$ for any Borel subset $A \subseteq Y$. Then for any Borel function $h : Y \to \mathbb{R}$ such that $h \in L^1(f_*(\mu))$,

$$\int_Y h \, df_*(\mu) = \int_X (h \circ f) \, d\mu.$$

*Moreover, the map f_* is **affine**: for any $r_1, \dots, r_n \in [0,1]$ with $\sum_{i=1}^n r_i = 1$ and $\mu_1, \dots, \mu_n \in \mathcal{M}(X)$,*

$$f_*\biggl(\sum_{i=1}^n r_i \mu_i\biggr) = \sum_{i=1}^n r_i f_*(\mu_i).$$

(iv) *If $f : X \to Y$ is continuous, then $f_* : \mathcal{M}(X) \to \mathcal{M}(Y)$ is continuous. If $f : X \to Y$ is also surjective, then f_* is also surjective.*

Remark 2.2. Let X be any compact Hausdorff space and let $C(X)$ be the collection of continuous functions from X to $\mathbb{R}$. We consider $C(X)$ as a normed vector space with the $\|\cdot\|_\infty$ norm, i.e., $\|f\|_\infty = \sup_{x \in X} |f(x)|$. The dual of $C(X)$, denoted by $C(X)^*$, is the space of all continuous linear functionals, i.e., maps from $C(X)$ to $\mathbb{R}$ which are continuous with respect to the norm topology on $C(X)$. The weak-* topology on $C(X)^*$ is the coarsest topology such that for any $a \in X$, the map $E_a : C(X) \to \mathbb{R}$ given by $E_a(f) = f(a)$ is continuous. We remark that $\mathcal{M}(X)$ can be naturally viewed as a subset of $C(X)^*$ via $\mu \mapsto \int - d\mu$. The topology induced from $C(X)^*$ on $\mathcal{M}(X)$ is both compact and Hausdorff. Moreover, $\mathcal{M}(X)$ forms a convex subset of $C(X)^*$.

Convention 2.3. If $f : X \to \mathbb{R}$ is a measurable function, we sometimes write $\int_X f\, d\mu$ simply as $\mu(f)$.

Definition 2.4. Let X be a compact Hausdorff space and $\mu \in \mathcal{M}(X)$. The *support of* μ is $\mathrm{supp}(\mu) := \{a \in X : \mu(U) > 0 \text{ for any open neighborhood } U \text{ of } a\}$. Then $\mathrm{supp}(\mu)$ is a nonempty closed subset of X. We remark that $\mu(\mathrm{supp}(\mu)) = 1$.

By a *compact group* we mean a compact Hausdorff topological group where both the multiplication $- \cdot - : G \times G \to G$ and inverse ${}^{-1} : G \to G$ maps are continuous.

Definition 2.5. Let G be a compact group and $\mu, \nu \in \mathcal{M}(G)$. Then their *convolution product*[1] $\mu \star \nu$ is the unique regular Borel measure on G such that for any continuous function $f : G \to \mathbb{R}$,

$$\int_G f(z)\, d(\mu \star \nu)(z) = \int_G \int_G f(x \cdot y)\, d\mu(x)\, d\nu(y).$$

Equivalently, $\mu \star \nu$ is the unique regular Borel measure on G such that for any Borel subset E of G,

$$\mu \star \nu(E) = \int_G \mu(Ex^{-1})\, d\nu(x).$$

See, e.g., [Stromberg 1959] for a proof the this equivalence.

Remark 2.6. Let G be a compact group.

(1) If $a, b \in G$, then $\delta_{a \cdot b} = \delta_a \star \delta_b$ (where δ_a denotes the Dirac measure on a).

[1] To stay consistent with the notation in [Chernikov and Gannon 2022], we will use "$*$" to denote definable convolution (defined later in this section) and "$\star$" to denote classical convolution.

(2) The space $\mathcal{M}(G)$ is a compact topological semigroup under convolution. In particular, the map $\star : \mathcal{M}(G) \times \mathcal{M}(G) \to \mathcal{M}(G)$ is associative and continuous.

(3) The map $\delta : G \to \mathcal{M}(G)$, $a \mapsto \delta_a$ is an embedding of topological semigroups.

Definition 2.7. Suppose that G is a compact group and $\lambda \in \mathcal{M}(G)$. We say that λ is *idempotent* if $\lambda \star \lambda = \lambda$.

The following theorem classifies idempotent measures on compact groups. The first proof of this theorem is due to Kawada and Itô [1940, Theorem 3] and uses representation theory of compact groups. This result was rediscovered a decade-and-a-half later by Wendel [1954, Theorem 1] using semigroup theory.

Fact 2.8. *Suppose G is a compact group and $\lambda \in \mathcal{M}(G)$. Then the following are equivalent*:

(1) *λ is idempotent.*

(2) $\mathrm{supp}(\lambda)$ *is a closed subgroup of G and $\lambda|_{\mathrm{supp}(\lambda)}$ is the normalized Haar measure on* $\mathrm{supp}(\lambda)$.

We are interested in which ways this theorem can be recovered for Keisler measures on definable groups. However, finding subgroups of a monster model is more difficult than directly applying this classification theorem since the support of an *idempotent Keisler measure* is a collection of types and not a subgroup of the model. Instead, we will also need to take into account a measure's stabilizer. This distinction does not arise in the compact group setting since the stabilizer of an idempotent probability measure is the same as its support. We take a moment to be precise about this statement.

Definition 2.9. Suppose G is a compact group and $\lambda \in \mathcal{M}(G)$. Its *right stabilizer* is $\mathrm{Stab}(\lambda) := \{a \in G : \lambda(B \cdot a) = \lambda(B)$ for any Borel set $B \subseteq G\}$.

Lemma 2.10. *Let G be a compact group and $\lambda \in \mathcal{M}(G)$. If λ is idempotent, then* $\mathrm{supp}(\lambda) = \mathrm{Stab}(\lambda)$.

Proof. Suppose $a \in \mathrm{supp}(\lambda)$. By Fact 2.8, $\mathrm{supp}(\lambda)$ is a closed subgroup of G and $\lambda|_{\mathrm{supp}(\lambda)}$ is the normalized Haar measure on $\mathrm{supp}(\lambda)$. Hence $\lambda(C \cdot a) = \lambda(C)$ for any Borel subset C of $\mathrm{supp}(\lambda)$. Let X be a Borel subset of G. Then

$$\lambda(X \cdot a) = \lambda\big((X \cdot a) \cap \mathrm{supp}(\lambda)\big) = \lambda\big((X \cap \mathrm{supp}(\lambda)) \cdot a\big) = \lambda(X \cap \mathrm{supp}(\lambda)) = \lambda(X),$$

and hence $a \in \mathrm{Stab}(\lambda)$.

Conversely, suppose $a \in \mathrm{Stab}(\lambda)$, but $a \notin \mathrm{supp}(\lambda)$. By Fact 2.8, this implies that $(\mathrm{supp}(\lambda) \cdot a) \cap \mathrm{supp}(\lambda) = \varnothing$. However $\lambda(\mathrm{supp}(\lambda)) = 1$ and also

$$\lambda(\mathrm{supp}(\lambda) \cdot a) = \lambda\big(\mathrm{supp}(\lambda) \cdot a \cap \mathrm{supp}(\lambda)\big) = \lambda(\varnothing) = 0,$$

a contradiction. □

Finally, we recall a couple of facts on integrating functions over compact groups.

Fact 2.11. *Suppose that G is a compact group and H is a closed subgroup of G with normalized Haar measure λ_H. Let $h \in H$, and let $f : G \to \mathbb{R}$ be a Borel function such that $f|_H \in L^1(\lambda_H)$, i.e., the restriction of f to H is integrable. Then*

$$\int_G f(x)\, d\lambda(x) = \int_G f(x \cdot h)\, d\lambda(x),$$

where λ is the measure on G defined by $\lambda(X) = \lambda_H(X \cap H)$.

The next fact appears hard to find explicitly stated in the literature, so we provide a proof for completeness.

Lemma 2.12. *Let G be a compact group and assume that $f : G \to \mathbb{R}$ is continuous. Let $\mu \in \mathcal{M}(G)$. Then the map $b \mapsto \int_G f(x \cdot b)\, d\mu$ from G to $\mathbb{R}$ is continuous.*

Proof. Define $h : G \to \mathbb{R}$ via $h(b) = \int_G f(x \cdot b)\, d\mu$. We first show that for every $\varepsilon > 0$ there exists an open neighborhood U of the identity $e \in G$ such that for any $b \in U$, $\sup_{x \in G} |f(x) - f(x \cdot b)| < \varepsilon$. Fix $\varepsilon > 0$, and suppose the statement does not hold. Then for every neighborhood U of e there exist some $b_U \in U$ and $c_U \in G$ such that $|f(c_U) - f(c_U \cdot b_U)| \geq \varepsilon$. Let N be the set of all open neighborhoods of e. Then N is a directed set under reverse inclusion and $(c_U \cdot b_U)_{U \in N}$ is a net. Since G is compact, we may pass to a convergent subnet N' of N so that $(c_U \cdot b_U)_{U \in N'}$ converges. Note also that still $\lim_{U \in N'} b_U = e$. Since the nets $(c_U \cdot b_U)_{U \in N'}$ and $(b_U)_{U \in N'}$ both converge and G is a topological group, the net $(c_U)_{U \in N'}$ also converges. Let $c := \lim_{U \in N'} c_U$. By continuity of f,

$$\lim_{U \in N'} f(c_U) = f(c) = \lim_{U \in N'} f(c_U \cdot b_U).$$

Then $\lim_{U \in N'} |f(c_U) - f(c_U \cdot b_U)| = 0$, but $|f(c_U) - f(c_U \cdot b_U)| \geq \varepsilon$ for each $U \in N'$ by assumption, a contradiction.

We now show that h is continuous. Let $(r, s) \subseteq \mathbb{R}$, $g_0 \in h^{-1}((r, s))$, and $\varepsilon := \min\{|h(g_0) - r|, |h(g_0) - s|\}$. By the paragraph above, there exists an open neighborhood of the identity U such that $\sup_{x \in G} |f(x) - f(x \cdot b)| < \varepsilon/2$ for any $b \in U$. We will show that the open set $g_0 \cdot U$ is a subset of $h^{-1}((r, s))$ containing g_0. Note that $g_0 \in g_0 \cdot U$ since $e \in U$. Now suppose that $g_1 \in g_0 \cdot U$, so that $g_1 = g_0 \cdot b_1$ for some $b_1 \in U$. Since, for any $g \in G$, the action $k(x) \mapsto k(x \cdot g)$ of G on the space $C(G)$ of continuous functions from G to $\mathbb{R}$ preserves the uniform norm, acting on the right by g_0 derives $\sup_{x \in G} |f(x \cdot g_0) - f(x \cdot g_0 \cdot b_1)| < \varepsilon/2$. Therefore

$$h(g_1) = \int_G f(x \cdot g_1)\, d\mu = \int_G f(x \cdot g_0 \cdot b_1)\, d\mu \approx_{\varepsilon/2} \int_G f(x \cdot g_0)\, d\mu = h(g_0).$$

Hence $h(g_1) \in (r, s)$ and thus $g_0 \cdot U$ is an open subset of $h^{-1}((r, s))$. Therefore $h^{-1}((r, s))$ is also open, and the map h is continuous. $\square$

2B. ***Model-theoretic setting.*** For the most part, our notation is standard. Let T be a complete first-order theory in a language $\mathcal{L}$ and assume that $\mathcal{U}$ is a sufficiently saturated and homogeneous model of T. While the rest of the paper is focused on the setting where T expands the theory of a group, this section contains results about arbitrary theories. We write $x, y, z, \ldots$ to denote arbitrary finite tuples of variables. If x is a tuple of variables and $A \subseteq \mathcal{U}$, then $\mathcal{L}_x(A)$ is the collection of formulas with free variables in x and parameters from A, modulo logical equivalence. We write $\mathcal{L}_x$ for $\mathcal{L}_x(\varnothing)$. Given a partitioned formula $\varphi(x; y)$ with object variables x and parameter variables y, we let $\varphi^*(y; x) := \varphi(x; y)$ be the partitioned formula with the roles of x and y reversed.

As usual, $S_x(A)$ denotes the space of types over A, and if $A \subseteq B \subseteq \mathcal{U}$ then $S_x^{\mathrm{fs}}(B, A)$ (respectively, $S_x^{\mathrm{inv}}(B, A)$) denotes the closed set of types in $S_x(B)$ that are finitely satisfiable in A (respectively, invariant over A). Throughout this paper, we will want to discuss the spaces $S_x^{\mathrm{inv}}(B, A)$ and $S_x^{\mathrm{fs}}(B, A)$ simultaneously, so we let $S_x^{\dagger}(B, A)$ denote "either $S_x^{\mathrm{fs}}(B, A)$ or $S_x^{\mathrm{inv}}(B, A)$". If $\varphi(x) \in \mathcal{L}_x(\mathcal{U})$, then $[\varphi(x)] = \{p \in S_x(\mathcal{U}) : \varphi(x) \in p\}$. Given a set $X \subset \mathcal{U}^x$ and $A \subseteq \mathcal{U}$ a small set of parameters, we say that X is $\bigvee$-definable over A (respectively, $\bigwedge$- or type-definable over A) if for some $\{\psi_i(x)\}_{i \in I}$ with $\psi_i(x) \in \mathcal{L}_x(A)$ we have $X = \bigcup_{i \in I} \psi_i(\mathcal{U})$ (respectively, $X = \bigcap_{i \in I} \psi_i(\mathcal{U})$). And X is $\bigvee$-definable (respectively, type-definable) if it is $\bigvee$-definable (respectively, type-definable) over A for some small $A \subseteq \mathcal{U}$.

Definition 2.13. If X is a $\bigvee$-definable subset of $\mathcal{U}^x$, we let $[X] := \bigcup_{i \in I} [\psi_i(x)]$ where $\bigvee_{i \in I} \psi_i(x)$ is any $\bigvee$-definition of X. Likewise, if X is a type-definable subset of $\mathcal{U}^x$, we let $[X] := \bigcap_{i \in I} [\phi_j(x)]$, where $\bigwedge_{i \in I} \phi_j(x)$ is any $\bigwedge$-definition of X. Note that $[X]$ does not depend on the choice of the small set of formulas defining X.

In the next fact, (1) follows by considering the preimages of half-open intervals, and for a proof of (2) see, e.g., [Gannon 2019, Fact 2.10].

Fact 2.14. *Let S be a topological space and $f : S \to \mathbb{R}$ a function.*

(1) *Assume f is bounded and Borel. Then for every $\varepsilon > 0$ there exist $r_1, \ldots, r_n \in \mathbb{R}$ and Borel sets $B_1, \ldots, B_n$ such that $\{B_i\}_{i=1}^n$ partition S and*

$$\sup_{a \in S} \left| f(a) - \sum_{i=1}^n r_i \mathbf{1}_{B_i}(a) \right| < \varepsilon.$$

(2) *Assume S is a Stone space and f is continuous. Then for every $\varepsilon > 0$ there exists clopen sets $C_1, \ldots, C_n \subseteq S$ and $r_1, \ldots, r_n \in \mathbb{R}$ such that*

$$\sup_{a \in S} \left| f(a) - \sum_{i=1}^n r_i \mathbf{1}_{C_i}(a) \right| < \varepsilon.$$

2C. *Keisler measures.* For any $A \subseteq \mathcal{U}$, a *Keisler measure* over A in variables x is a finitely additive probability measure on $\mathcal{L}_x(A)$. We denote the space of Keisler measures over A (in variables x) as $\mathfrak{M}_x(A)$. Every $\mu \in \mathfrak{M}_x(A)$ extends uniquely to a regular Borel probability measure $\tilde{\mu}$ on the space $S_x(A)$, and we will routinely use this correspondence. If $A \subseteq B \subseteq \mathcal{U}$, then there is an obvious restriction map $r_0 : \mathfrak{M}_x(B) \to \mathfrak{M}_x(A)$ and we denote $r_0(\mu)$ simply as $\mu|_A$. Conversely, every $\mu \in \mathfrak{M}_x(A)$ admits an extension to some $\mu' \in \mathfrak{M}_x(B)$ (not necessarily a unique one).

Definition 2.15. Let $B \subseteq \mathcal{U}$ and $\mu \in \mathfrak{M}_x(\mathcal{U})$. We say that μ is

(1) *invariant over B* if for any formula $\varphi(x, y) \in \mathcal{L}_{x,y}(B)$ and elements $a, b \in \mathcal{U}^y$ such that $a \equiv_B b$ we have $\mu(\varphi(x, b)) = \mu(\varphi(x, a))$;

(2) *finitely satisfiable in B* if for any formula $\varphi(x) \in \mathcal{L}_x(\mathcal{U})$ such that $\mu(\varphi(x)) > 0$, there exists some $b \in B$ such that $\models \varphi(b)$.

We let $\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{U}, B)$ (respectively, $\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{U}, B)$) denote the closed set of Keisler measures in $\mathfrak{M}_x(\mathcal{U})$ that are finitely satisfiable in B (respectively, invariant over B).

Just as with types, we let $\mathfrak{M}^{\dagger}_x(\mathcal{U}, B)$ mean "$\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{U}, B)$ or $\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{U}, B)$". The *support of* $\mu \in \mathfrak{M}_x(B)$ is the nonempty closed set of types

$$\sup(\mu) = \{p \in S_x(B) : \mu(\varphi(x)) > 0 \text{ for any } \varphi(x) \in p\}.$$

Given $\bar{p} = (p_1, \dots, p_n)$ with $p_i \in S_x(A)$, we let $\mathrm{Av}(\bar{p}) \in \mathfrak{M}_x(A)$ be defined by $\mathrm{Av}(\bar{p})(\varphi(x)) := |\{i \in [n] : \varphi(x) \in p_i\}|/n$, and given $\bar{a} = (a_1, \dots, a_n) \in \mathcal{U}^x$, we let $\mathrm{Av}(\bar{a}) := \mathrm{Av}\big(\mathrm{tp}(a_1/\mathcal{U}), \dots, \mathrm{tp}(a_n/\mathcal{U})\big)$. We refer to, e.g., [Chernikov and Gannon 2022, Section 2] for a more detailed discussion of the aforementioned notions.

Definition 2.16. Let $\mathbb{X} \subseteq S_x(\mathcal{U})$. We let $\mathfrak{M}(\mathbb{X}) := \{\mu \in \mathfrak{M}_x(\mathcal{U}) : \sup(\mu) \subseteq \mathbb{X}\}$ be the set of Keisler measures supported on $\mathbb{X}$. If $\mathbb{X}$ is a closed subset of $S_x(\mathcal{U})$, we let $\mathcal{M}(\mathbb{X})$ denote the set of regular Borel probability measures on $\mathbb{X}$, with the topology on $\mathbb{X}$ induced from $S_x(\mathcal{U})$. When we consider $\mathcal{M}(\mathbb{X})$ as a topological space, we will always consider it with the weak-* topology.

The space of Keisler measures $\mathfrak{M}_x(A)$ is a (closed convex) subset of a real locally convex topological vector space of bounded charges on $\mathcal{L}_x(A)$ (see, e.g., [Bhaskara Rao and Bhaskara Rao 1983] for the details).

Lemma 2.17. *Assume that $\mathbb{X}$ is a closed subset of $S_x(\mathcal{U})$. Then $\mathfrak{M}(\mathbb{X})$ is a closed convex subset of $\mathfrak{M}_x(\mathcal{U})$.*

Proof. Suppose $\mathfrak{M}(\mathbb{X})$ is not closed. Then $\lim_{i \in I} \mu_i = \mu$ for some $\mu \notin \mathfrak{M}(\mathbb{X})$ and some net $(\mu_i)_{i \in I}$ with $\mu_i \in \mathfrak{M}(\mathbb{X})$. Then there exists a type $p \in \sup(\mu) \setminus \mathbb{X}$. Since $\mathbb{X}$ is closed, the set $U := S_x(\mathcal{U}) \setminus \mathbb{X}$ is open. Hence $U = \bigcup_{j \in J} [\varphi_j(x)]$ for some set of formulas $\{\varphi_j\}_{j \in J}$ and there is some $j \in J$ such that $\varphi_j(x) \in p$. Then $[\varphi_j(x)] \cap \mathbb{X} = \varnothing$ and $\mu(\varphi_j(x)) > 0$ (since $p \in \sup(\mu)$). Thus $\lim_{i \in I} \mu_i(\varphi_j(x)) = \lim_{i \in I} 0 = 0 < \mu(\varphi_j(x))$, a contradiction.

The space $\mathfrak{M}(\mathbb{X})$ is convex since if $r, s \in \mathbb{R}_{>0}$ with $r + s = 1$, and $\mu, \nu \in \mathfrak{M}(\mathbb{X})$, then $\sup(r\mu + s\nu) = \sup(\mu) \cup \sup(\nu) \subseteq \mathbb{X}$. $\square$

In the later sections, we will need to discuss maps from the space of Keisler measures to other spaces of measures. The following definition is an appropriate notion of an isomorphism in this context (and will be denoted by $\cong$).

Definition 2.18. Let V_1, V_2 be two locally convex topological vector spaces. Suppose that C_1 and C_2 are closed convex subsets of V_1 and V_2, respectively. A map $f : C_1 \to C_2$ is an *affine homeomorphism* if f is a homeomorphism from C_1 to C_2 (with the induced topologies) and for any $a_1, \dots, a_n \in C_1$ and $r_1, \dots, r_n \in \mathbb{R}_{\geq 0}$ with $\sum_{i=1}^{n} r_i = 1$ we have

$$f\left(\sum_{i=1}^{n} r_i a_i\right) = \sum_{i=1}^{n} r_i f(a_i).$$

Definition 2.19. Let A be a subset of a locally convex topological vector space, V, and let $b \in V$. We say that b is *extreme in* A (or an *extreme point of* A) if $b \in A$ and b cannot be written as $rc_1 + (1-r)c_2$ for $c_1, c_2 \in A$ where $c_1 \neq c_2$ and $r \in (0, 1)$. We let $\mathrm{ex}(A) := \{c \in A : c \text{ is extreme in } A\}$.

Fact 2.20 (Krein–Milman theorem). *Let A be a convex compact subset of a locally convex topological vector space V. Then the convex hull of* $\mathrm{ex}(A)$ *is a dense subset of A.*

Proposition 2.21. *Let $\mathbb{X} \subseteq S_x(\mathcal{U})$ be a closed set. Then there exists an affine homeomorphism $\gamma : \mathfrak{M}(\mathbb{X}) \to \mathcal{M}(\mathbb{X})$ such that for any $\varphi(x) \in \mathcal{L}_x(\mathcal{U})$ and $\mu \in \mathfrak{M}(\mathbb{X})$,*

$$\mu(\varphi(x)) = \gamma(\mu)([\varphi(x)] \cap \mathbb{X}).$$

Moreover, $\sup(\mu) = \mathrm{supp}(\gamma(\mu))$.

Proof. This follows directly from the fact that every Keisler measure μ in $\mathfrak{M}_x(\mathcal{U})$ extends uniquely to a regular Borel probability measure $\tilde{\mu}$ on $S_x(\mathcal{U})$. We let $\gamma(\mu) := \tilde{\mu} \restriction_{\mathbb{X}}$, i.e., the restriction of the measure $\tilde{\mu}$ to the collection of Borel subsets of $\mathbb{X}$. See, e.g., [Simon 2015, page 99] for the details. $\square$

For a proof of the following fact, see [Chernikov and Gannon 2022, Lemma 2.10].

Fact 2.22. (1) $\mu \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{U}, M)$ *if and only if* $p \in S_x^{\mathrm{fs}}(\mathcal{U}, M)$ *for all* $p \in \sup(\mu)$.

(2) (T is NIP) $\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{U}, M)$ *if and only if* $p \in S_x^{\mathrm{inv}}(\mathcal{U}, M)$ *for every* $p \in \sup(\mu)$.

Combining Proposition 2.21 and Fact 2.22 we have the following.

Corollary 2.23. (1) *If T is any theory, then* $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{U}, M) = \mathfrak{M}(S_x^{\mathrm{fs}}(\mathcal{U}, M)) \cong \mathcal{M}(S_x^{\mathrm{fs}}(\mathcal{U}, M))$.

(2) *If T is NIP, then* $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{U}, M) = \mathfrak{M}(S_x^{\mathrm{inv}}(\mathcal{U}, M)) \cong \mathcal{M}(S_x^{\mathrm{inv}}(\mathcal{U}, M))$.

Remark 2.24. It is not true that $\mathfrak{M}(S_x^{\mathrm{inv}}(\mathcal{U}, M)) = \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{U}, M)$ in an arbitrary theory; see [Chernikov and Gannon 2022, Lemma 2.10(4)].

Lemma 2.25. *For any $\mu \in \mathfrak{M}_x(\mathcal{U})$, there exists a net of measures $(\nu_j)_{j\in J}$ in $\mathfrak{M}_x(\mathcal{U})$ such that*

(1) *for each $j \in J$, $\nu_j = \mathrm{Av}(\bar{p}_j)$ for some $\bar{p}_j = (p_{j_1}, \ldots, p_{j_m})$ with $p_{j_i} \in \sup(\mu)$;*
(2) $\lim_{j\in J} \nu_j = \mu$.

Moreover, if μ is finitely satisfiable in $M \preceq \mathcal{U}$, then we can take ν_j of the form $\mathrm{Av}(\bar{a}_j)$ for some $\bar{a}_j \in (M^x)^{<\omega}$.

Proof. Consider a basic open subset O of $\mathfrak{M}_x(\mathcal{U})$, of the form

$$O = \bigcap_{i=1}^{n} \{\nu \in \mathfrak{M}_x(\mathcal{U}) : r_i < \nu(\theta_i(x)) < s_i\}.$$

Suppose that $\mu \in O$. Let $\mathcal{B}$ be the (finite) Boolean algebra generated by the sets $\{\theta_1(x), \ldots, \theta_n(x)\}$, and let $\{\sigma_1(x), \ldots, \sigma_m(x)\}$ be the set of its atoms. For each atom $\sigma_i(x)$ such that $\mu(\sigma_i(x)) > 0$, there exists some $p_i \in \sup(\mu)$ such that $\sigma_i(x) \in p_i$. Consider the measure

$$\lambda := \sum_{\{i\in[n]:\mu(\sigma_i(x))>0\}} \mu(\sigma_i(x))\delta_{p_i}.$$

Then $\lambda(\theta_i(x)) = \mu(\theta_i(x))$ for every $i \in [n]$, and hence $\lambda \in O$. We can choose a sufficiently large $t \in \mathbb{N}$ and $s_i \in \mathbb{N}$ so that s_i/t is sufficiently close to $\mu(\sigma_i(x))$, so that $\nu_O := \sum_{\{i\in[n]:\mu(\sigma_i(x))>0\}} (s_i/t)\delta_{p_i} \in O$ (taking $\bar{p}_O$ to be the tuple of types of length t with p_i repeated s_i times, we see that $\nu_O = \mathrm{Av}(\bar{p}_O)$). Then we can take the net $(\nu_O)_{\mu\in O}$.

And if μ is finitely satisfiable in M and $\mu(\sigma_i(x)) > 0$, then $\models \sigma_i(a_i)$ for some $a_i \in M^x$, and we can take $p_i := \mathrm{tp}(a_i/\mathcal{U})$ (see [Chernikov and Gannon 2022, Proposition 2.11]). □

2D. ***Definable convolution in NIP groups.*** In this section, we assume that T is an $\mathcal{L}$-theory expanding a group, we denote by $\mathcal{G}$ a sufficiently saturated model of T and by G a small elementary submodel; $x, y, \ldots$ denote singleton variables; and for any $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$, we let $\varphi'(x, y) := \varphi(x \cdot y)$.

Definition 2.26 (T is NIP). Suppose that $\mu, \nu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$. Then we define $\mu * \nu$ to be the unique Keisler measure in $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ such that for any formula $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$,

$$\mu * \nu(\varphi(x)) = \mu_x \otimes \nu_y(\varphi(x \cdot y)) = \int_{S_y(G')} F_{\mu,G'}^{\varphi'}\, d(\nu_{G'}),$$

where G' is a small model containing G and all parameters from φ, the map $F_{\mu,G'}^{\varphi'} : S_y(G') \to [0, 1]$ is given by $F_{\mu,G'}^{\varphi'}(q) = \mu(\varphi(x \cdot b))$ for some (equivalently, any)

$b \in \mathcal{G}$ with $b \models q$, and $\nu_{G'}$ is the regular Borel probability measure on $S_y(G')$ corresponding to the Keisler measure $\nu|_{G'}$. We will routinely suppress notation and write this integral as $\int_{S_y(G')} F_\mu^{\varphi'}\, d\nu$.

Remark 2.27. This integral is well defined since invariant measures in NIP are Borel-definable, so the maps which are being integrated are measurable, and does not depend on the choice of G'. For more details about definable convolution and its basic properties we refer the reader to [Chernikov and Gannon 2022, Section 3.2]. In particular, we will freely use [Chernikov and Gannon 2022, Proposition 3.14].

The following is well known; see, e.g., [Chernikov and Gannon 2022, Fact 3.11].

Fact 2.28. *Both* $(S_x^{\mathrm{inv}}(\mathcal{G}, G), *)$ *and* $(S_x^{\mathrm{fs}}(\mathcal{G}, G), *)$ *are left continuous* (i.e., $p \mapsto p * q$ is a continuous map for every q) *compact Hausdorff semigroups.*

The next fact is from [Chernikov and Gannon 2022, Propositions 6.2(3) and 6.4].

Fact 2.29 (T is NIP). *Both* $(\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G), *)$ *and* $(\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G), *)$ *are left continuous* (i.e., $\mu \mapsto \mu * \nu$ is a continuous map for every ν) *compact Hausdorff semigroups.*

Moreover, for any fixed ν *and* $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$, *the map* $\mu \mapsto (\mu * \nu)(\varphi(x)) \in [0, 1]$ *is continuous.*

We also have right continuity when multiplying by a *definable* measure (but not in general).

Lemma 2.30. *If* $\nu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ *is a definable measure, then the map* $\mu \mapsto \nu * \mu$ *from* $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ *to* $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ *is continuous.*

Proof. Let O be a basic open subset of $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$, that is,

$$O = \bigcap_{i=1}^{n} \{\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G) : r_i < \mu(\varphi_i(x)) < s_i\}$$

for some $r_i, s_i \in \mathbb{R}$ and $\varphi_i(x) \in \mathcal{L}_x(\mathcal{G})$. We have

$$\begin{aligned}
(\nu * -)^{-1}(O) &= \bigcap_{i=1}^{n} \{\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G) : r_i < (\nu * \mu)(\varphi_i(x)) < s_i\} \\
&= \bigcap_{i=1}^{n} \{\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G) : r_i < (\nu_z \otimes \mu_x)(\varphi_i(z \cdot x)) < s_i\} \\
&= \bigcap_{i=1}^{n} \Big(\big((\nu_z \otimes -)(\varphi_i(z \cdot x))\big)^{-1}(r_i, s_i)\Big),
\end{aligned}$$

where ν_z is simply ν_x with change of variables to z and (r_i, s_i) is an open subinterval of $[0, 1]$. By, e.g., [Conant et al. 2021, Lemma 5.4], the map $\mu_x \in \mathfrak{M}_x(\mathcal{G}) \mapsto (\nu_z \otimes \mu_x)(\varphi_i(z \cdot x)) \in [0, 1]$ is continuous, so its restriction to $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ remains continuous. Thus O is open, as the intersection of finitely many open sets. $\square$

Definition 2.31. A measure $\mu \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ is *idempotent* if $\mu * \mu = \mu$.

The following simple observation will be frequently used in computations.

Fact 2.32. *Let $\mu \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ and $f : S_x(G) \to \mathbb{R}$ be a bounded Borel function. Let $r : S_x(\mathcal{G}) \to S_x(G)$, $p \mapsto p|_G$ be the restriction map. Then*

$$\int_{S_x(G)} f \, d\mu_G = \int_{S_x(\mathcal{G})} (f \circ r) \, d\mu.$$

2E. *Some facts from Ellis semigroup theory.*

Definition 2.33. Suppose that $(X, *)$ is a semigroup. A nonempty subset I of X is a *left ideal* if $XI = \{x * i : x \in X, i \in I\} \subseteq I$. We say that I is a *minimal left ideal* if I does not properly contain any other left ideal.

The next fact summarizes the results that we will need from the theory of Ellis semigroups. See [Ellis et al. 2001, Proposition 4.2; Glasner 2007, Proposition 2.4].

Fact 2.34. *Suppose that X is a compact Hausdorff space and $(X, *)$ is a left continuous semigroup, i.e., for each $q \in X$, the map $- * q : X \to X$ is continuous. Then there exists a minimal left ideal I, and any minimal left ideal is closed. We let* $\mathrm{id}(I) = \{u \in I : u^2 = u\}$ *be the set of idempotents in I.*

(1) $\mathrm{id}(I)$ *is nonempty.*

(2) *For every $p \in I$ and $u \in \mathrm{id}(I)$, $p * u = p$.*

(3) *For every $u \in \mathrm{id}(I)$, $u * I = \{u * p : p \in I\} = \{p \in I : u * p = p\}$ is a subgroup of I with identity element u. For every $u' \in \mathrm{id}(I)$, the map $\rho_{u,u'} := (u' * -)|_{u*I}$ is a group isomorphism from $u * I$ to $u' * I$. In view of this, we refer to $u * I$ as* ***the ideal group****.*

(4) $I = \bigcup\{u * I : u \in \mathrm{id}(I)\}$, *where the sets in the union are pairwise disjoint, and each set $u \cdot I$ is a subgroup of I with identity u.*

(5) *For any $q \in X$, $I * q$ is a minimal left ideal; and if $p \in I$, then $X * p = I$.*

(6) *Let J be another minimal left ideal of X and $u \in \mathrm{id}(I)$. Then there exists a unique idempotent $u' \in \mathrm{id}(J)$ such that $u * u' = u'$ and $u' * u = u$. The map $\rho_{I,J} := (- * u')|_I$ is a homeomorphism from I to J (with the induced topologies) mapping $u * I$ to $u' * J$.*

The following is a celebrated theorem of Ellis [1957, Theorems 1 and 2] (see also [Lawson 1974, Corollary 5.2]).

Fact 2.35 (Ellis joint continuity theorem). (1) *Let G be a locally compact Hausdorff semitopological group* (i.e., G is equipped with a group structure such that the maps $x \mapsto y \cdot x$ and $x \mapsto x \cdot y$ from G into G are continuous for any fixed $y \in G$), *and let X be a locally compact Hausdorff topological space. Then every separately continuous action of G on X is (jointly) continuous.*

(2) *If G is a locally compact Hausdorff semitopological group, then G is a topological group.*

2F. *Some facts from Choquet theory.* We recall some notions and facts from Choquet theory for *not necessarily metrizable* compact Hausdorff spaces (we use [Phelps 2001] as a general reference). Let E be a locally convex real topological vector space. The following generalizes the usual notion of a simplex in $\mathbb{R}^n$ to the infinite-dimensional context.

Definition 2.36 [Phelps 2001, Section 10]. (1) A set $P \subseteq E$ is a *convex cone* if $P + P \subseteq P$ and $\lambda P \subseteq P$ for every scalar $\lambda > 0$ in $\mathbb{R}$.

(2) A set $X \subseteq P$ is the *base* of a convex cone P if for every $y \in P$ there exists a unique scalar $\lambda \geq 0$ in $\mathbb{R}$ and $x \in X$ such that $y = \lambda x$ (not all convex cones have a base).

(3) A convex cone P in E induces a translation-invariant partial ordering on E: $x \geq y$ if and only if $x - y \in P$. When P admits a base, $P \cap (-P) = \{0\}$, and hence $x \geq y \wedge y \geq x \implies x = y$.

(4) A nonempty compact convex set $X \subseteq E$ is a *Choquet simplex*, or just *simplex*, if X is the base of a convex cone $P \subseteq E$ such that P is a lattice with respect to the ordering induced by P. That is, for every $x, y \in P$ there exists a greatest lower bound $z \in P$ (i.e., $z \leq x$ and $z \leq y$, and for every $z' \in P$ with $z' \leq x$ and $z' \leq y$, $z' \leq z$). The greatest lower bound z of x and y is unique and denoted by $x \wedge y$.

We could not find a direct quote for the following fact, so we provide a short argument combining several standard results in the literature.

Fact 2.37. *Let S be a compact Hausdorff space and $\mathcal{T}$ a family of continuous functions from S into S. Then the set of all regular $\mathcal{T}$-invariant* (that is, $\mu(T^{-1}(A)) = \mu(A)$ for every Borel $A \subseteq S$ and $T \in \mathcal{T}$) *Borel probability measures on S, denoted by $\mathcal{M}_{\mathcal{T}}(S)$, is a Choquet simplex* (*assuming it is nonempty*).

Proof. By the Riesz representation theorem, we can view the set $\mathcal{M}^+(S)$ of all regular Borel nonnegative finite measures on S as a subset of $C(S)^*$, the dual (real topological vector) space of the topological vector space of continuous functions on S, with the weak-* topology. Let $\mathcal{M}_{\mathcal{T}}(S)$ (respectively, $\mathcal{M}^+_{\mathcal{T}}(S)$) be the set of regular Borel $\mathcal{T}$-invariant probability (respectively, finite nonnegative) measures on S. Then $\mathcal{M}_{\mathcal{T}}(S) \subseteq \mathcal{M}^+_{\mathcal{T}}(S) \subseteq \mathcal{M}^+(S)$ are compact convex subsets (by Borel measurability of the maps in $\mathcal{T}$; see [Phelps 2001, page 76]). Moreover, $\mathcal{M}^+_{\mathcal{T}}(S)$ is a convex cone with the base $\mathcal{M}_{\mathcal{T}}(S)$. It is well known that $\mathcal{M}^+(S)$ forms a lattice: for $\mu, \nu \in \mathcal{M}^+(S)$, their greatest lower bound $\mu \wedge \nu \in \mathcal{M}^+(S)$ can be defined via $(\mu \wedge \nu)(A) = \inf_{B \in \mathcal{S}, B \subseteq A}\{\mu(B) + \nu(A \setminus B)\}$ (see, e.g., [Dales et al. 2016, page 111]; it is easy to verify from this definition that if μ, ν are regular, then $\mu \wedge \nu$ is also regular). Finally, [Phelps 2001, Proposition 12.3] shows that if $\mu, \nu \in \mathcal{M}^+(S)$ are $\mathcal{T}$-invariant, then $\mu \wedge \nu$ is also $\mathcal{T}$-invariant (using an equivalent definition of the lower bound in terms of the Radon–Nikodym derivative). Hence $\mathcal{M}^+_{\mathcal{T}}(S)$ is a lattice, and so $\mathcal{M}_{\mathcal{T}}(S)$ is a Choquet simplex. □

Definition 2.38 (see [Phelps 2001, Section 11] or [Alfsen 1971, Chapter 2, §4]). A compact convex set $X \subseteq E$ is a *Bauer simplex* if X is a Choquet simplex and $\mathrm{ex}(X)$ is closed.

Definition 2.39. A point $x \in E$ is the *barycenter* of a regular Borel probability measure μ on X if $f(x) = \mu(f) := \int_X f\, d\mu$ for any continuous linear function $f : E \to \mathbb{R}$.

Remark 2.40. Both the property of being a Choquet simplex and the property of being a Bauer simplex are preserved under affine homeomorphisms (see, e.g., [Phelps 2001, pages 52–53]).

Fact 2.41. (1) [Phelps 2001, Proposition 11.1] *X is a Bauer simplex if and only if the map sending a regular Borel probability measure μ on* $\overline{\mathrm{ex}(X)}$ (the closure of the extreme points) *to its barycenter is an affine homeomorphism of* $\mathcal{M}(\overline{\mathrm{ex}(X)})$ *and X* (*and thus a posteriori of* $\mathcal{M}(\mathrm{ex}(X))$ *and X*).

(2) [Alfsen 1971, Corollary II.4.4] *Up to affine homeomorphisms*, *Bauer simplices are exactly the sets of the form $\mathcal{M}(X)$ for X a compact Hausdorff space* (*where* $\mathrm{ex}(\mathcal{M}(X)) = \{\delta_x : x \in X\}$).

3. Definable convolution on $\mathcal{G}$ and convolution on $\mathcal{G}/\mathcal{G}^{00}$

Throughout the rest of the paper, T is a complete NIP theory expanding a group, $\mathcal{G}$ is a monster model of T, G is a small elementary submodel of $\mathcal{G}$, $x, y, \ldots$ denote singleton variables, and for any $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$, $\varphi'(x, y) = \varphi(x \cdot y)$. We define and study a natural pushforward map from $\mathfrak{M}_x(\mathcal{G})$ to $\mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$. We demonstrate that this map is a homomorphism from the semigroup $(\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G), *)$ of invariant Keisler measures with definable convolution onto the semigroup $(\mathcal{M}(\mathcal{G}/\mathcal{G}^{00}), \star)$ of regular Borel probability measures on the compact group $\mathcal{G}/\mathcal{G}^{00}$ with classical convolution. In particular, the image of an idempotent, invariant Keisler measure on $\mathcal{G}$ is an idempotent measure on the compact group $\mathcal{G}/\mathcal{G}^{00}$. The proofs of these theorems are primarily analytic, and the NIP assumption is used to ensure that $\mathcal{G}^{00}$ exists and definable convolution is well defined. We begin by recalling some properties of $\mathcal{G}/\mathcal{G}^{00}$ and define the corresponding pushforward map.

Fact 3.1. *Suppose that T is NIP.*

(i) *There exists a smallest type-definable subgroup of $\mathcal{G}$ of bounded index*, *denoted by $\mathcal{G}^{00}$. Moreover*, *$\mathcal{G}^{00}$ is a normal subgroup of $\mathcal{G}$ type-definable over $\varnothing$. Let $\pi : \mathcal{G} \to \mathcal{G}/\mathcal{G}^{00}$ be the quotient map, i.e.*, *$\pi(a) = a\mathcal{G}^{00}$.*

(ii) *$\mathcal{G}/\mathcal{G}^{00}$ is a compact group with the* ***logic topology***: *a subset B of $\mathcal{G}/\mathcal{G}^{00}$ is closed if and only if $\pi^{-1}(B)$ is type-definable over some/any small submodel of $\mathcal{G}$.*

(iii) *The map $\pi : \mathcal{G} \to \mathcal{G}/\mathcal{G}^{00}$ induces a continuous map $\hat{\pi} : S_x(\mathcal{G}) \to \mathcal{G}/\mathcal{G}^{00}$ via $\hat{\pi}(q) := \pi(a)$, where $a \models q|_G$ and G is some/any elementary submodel of $\mathcal{G}$. Therefore, we can consider the pushforward $\hat{\pi}_* : \mathcal{M}(S_x(\mathcal{G})) \to \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$. By Proposition 2.21, $\mathfrak{M}_x(\mathcal{G})$ is affinely homeomorphic to $\mathcal{M}(S_x(\mathcal{G}))$ and so (formally) we let $\pi_* : \mathfrak{M}_x(\mathcal{G}) \to \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ be the composition of $\hat{\pi}_*$ and this homeomorphism. We will primarily work with π_*, and usually identify $\hat{\pi}_*$ and π_* without comment.*

(iv) *The map $\pi_* : \mathfrak{M}_x(\mathcal{G}) \to \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ is continuous, affine, and surjective.*

Proof. (i) This is a theorem of Shelah [2008].

(ii) This is from [Pillay 2004] (see also [Simon 2015, Section 8]).

(iii) First, $\hat{\pi}$ is well defined. Indeed, let $G_1, G_2 \prec \mathcal{G}$ be small elementary submodels and $q \in S_x(\mathcal{G})$ be such that $a_i \models q|_{G_i}$ for $i \in \{1, 2\}$. It suffices to show $\pi(a_1) = \pi(a_2)$. Let U be an open subset of $\mathcal{G}/\mathcal{G}^{00}$ such that $\pi(a_1) \in U$, and we show that then also $\pi(a_2) \in U$. Since U is open, $\pi^{-1}(U)$ is $\bigvee$-definable over both G_1 and G_2. Let $\bigvee_{j\in I_i} \psi_j^i(x)$ be a definition of $\pi^{-1}(U)$ over G_i. Hence there is some $j_1 \in I_1$ such that $\mathcal{U} \models \psi_{j_1}(a_1)$, so $\psi_{j_1}(x) \in q$. As $\bigcup_{j\in I_1}[\psi_j^1(x)] = \bigcup_{j\in I_2}[\psi_j^2(x)]$ (see Definition 2.13), there exists some $j_2 \in I_2$ so that $\psi_{j_2}(x) \in q$. Now

$$a_2 \in \psi_{j_2}^2(\mathcal{U}) \subseteq \bigcup_{j\in I_2} \psi_j^2(\mathcal{U}) = \pi^{-1}(U) \implies \pi(a_2) \in U.$$

Since $\mathcal{G}/\mathcal{G}^{00}$ is Hausdorff and $\pi(a_1)$ and $\pi(a_2)$ are in the same open sets, we conclude that $\pi(a_1) = \pi(a_2)$.

By the previous paragraph, $\hat{\pi} = f \circ r_G$, where G is any small submodel, the map $r_G : S_x(\mathcal{G}) \to S_x(G)$ is the restriction map, and $f : S_x(G) \to \mathcal{G}/\mathcal{G}^{00}$ is defined via $f(q) = \pi(a)$, where $a \models q$. Both f and r_G are continuous maps and so $\hat{\pi}$ is a continuous map (the map f is continuous by (ii)).

(iv) This is by Fact 2.1(iii),(iv) and Proposition 2.21. □

Definition 3.2. We let $\pi_{G,*}^{\mathrm{fs}} := \pi_* \restriction_{\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G},G)}$ and $\pi_{G,*}^{\mathrm{inv}} := \pi_* \restriction_{\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G},G)}$. We will typically write $\pi_{G,*}^{\mathrm{inv}}$ simply as π_*^{inv} when G is clear from the context, and $\pi_*^{\dagger}$ to mean "either π_*^{inv} or π_*^{fs}".

Remark 3.3. Both π_*^{inv} and π_*^{fs} are continuous and affine since these maps are restrictions of π_* to a closed convex subspace.

Proposition 3.4. *The map $\pi_*^{\dagger} : \mathfrak{M}_x^{\dagger}(\mathcal{G}, G) \to \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ is surjective.*

Proof. Since $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G) \subseteq \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$, it suffices to show that π_*^{fs} is surjective. Fix $\nu \in \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$. By the Krein–Milman theorem, the convex hull of the extreme points of $\mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ is dense inside $\mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$. The extreme points of $\mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ are the Dirac measures concentrating on the elements of $\mathcal{G}/\mathcal{G}^{00}$ (see, e.g., [Simon 2011, Example 8.16]). Thus there exists a net $(\nu_i)_{i\in I}$ of measures in $\mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ such that $\lim_{i\in I}\nu_i=\nu$ and for each $i\in I$, $\nu_i=\sum_{j=1}^{n_i} r^i_j\delta_{b^i_j}$ for some $n_i\in\mathbb{N}$, $b^i_j\in\mathcal{G}/\mathcal{G}^{00}$ and $r^i_j\in\mathbb{R}_{>0}$ with $\sum_{j=1}^{n_i} r^i_j=1$. Since the map π is surjective, for each b^i_j there exists some $a^i_j\in\mathcal{G}$ such that $\pi(a^i_j)=b^i_j$. Let $p^i_j\in S^{\mathrm{fs}}_x(\mathcal{G},G)$ be a global coheir of $\mathrm{tp}(a^i_j/G)$, and let $\mu_i:=\sum_{j=1}^{n_i} r^i_j\delta_{p^i_j}$. Then $\pi_*(\mu_i)=\nu_i$. Now $(\mu_i)_{i\in I}$ is a net in the compact space $\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G},G)$, so, passing to a subnet, we may assume that it converges and let $\mu:=\lim_{i\in I}\mu_i$. Then

$$\pi_*(\mu)=\pi_*(\lim_{i\in I}\mu_i)=\lim_{i\in I}\pi_*(\mu_i)=\lim_{i\in I}\nu_i=\nu,$$

where the second equality follows from continuity of π_*. Hence π^{fs}_* is surjective. □

Lemma 3.5. *Let $p,q\in S^{\mathrm{inv}}_x(\mathcal{G},G)$. Then*

(i) $\hat{\pi}(p)\cdot\hat{\pi}(q)=\hat{\pi}(p*q)$,

(ii) $\pi_*(\delta_p)=\delta_{\hat{\pi}(p)}$,

(iii) $\pi_*(\delta_p*\delta_q)=\pi_*(\delta_p)\star\pi_*(\delta_q)$.

Proof. (i) Let $b\models q|_G$ and $a\models p|_{Gb}$. By definition $(a\cdot b)\models p*q|_G$, and hence

$$\hat{\pi}(p*q)=\pi(a\cdot b)=\pi(a)\cdot\pi(b)=\hat{\pi}(p)\cdot\hat{\pi}(q).$$

(ii) Let $f:\mathcal{G}/\mathcal{G}^{00}\to\mathbb{R}$ be a continuous function. Then

$$\pi_*(\delta_p)(f)=\int(f\circ\hat{\pi})\,d\delta_p=f(\hat{\pi}(p))=\int f\,d\delta_{\hat{\pi}(p)}=\delta_{\hat{\pi}(p)}(f).$$

Since $\pi_*(\delta_p)$ and $\delta_{\hat{\pi}(p)}$ agree on all continuous functions, by Fact 2.1(i) they belong to the same open sets in a Hausdorff space, so $\pi_*(\delta_p)=\delta_{\hat{\pi}(p)}$.

(iii) We have

$$\pi_*(\delta_p*\delta_q)=\pi_*(\delta_{p*q})=\delta_{\hat{\pi}(p*q)}=\delta_{\hat{\pi}(p)\cdot\hat{\pi}(q)}=\delta_{\hat{\pi}(p)}\star\delta_{\hat{\pi}(q)}.$$

Here the first equality follows from [Chernikov and Gannon 2022, Proposition 3.12], the second and third equalities follow from (ii) and (i) respectively, and the last equality is by Remark 2.6. □

To show that π^{inv}_* is a homomorphism, we first observe some basic properties of the action of $\mathcal{G}$ on its space of types and, in turn, on the space of continuous functions from $S_x(\mathcal{G})$ to $\mathbb{R}$.

Definition 3.6. Let G be a model of T. For $a\in G$ and $p\in S_x(G)$, let $p\cdot a:=\{\varphi(x\cdot a^{-1}):\varphi(x)\in p\}\in S_x(G)$ and $a\cdot p=\{\varphi(a^{-1}\cdot x):\varphi(x)\in p\}\in S_x(G)$. This defines a right (respectively, left) action of G on $S_x(G)$ by homeomorphisms.

Lemma 3.7. *For any $a \in \mathcal{G}$ and $q \in S_x(\mathcal{G})$ we have $\pi(a) \cdot \hat{\pi}(p) = \hat{\pi}(a \cdot p)$ and $\hat{\pi}(p) \cdot \pi(a) = \hat{\pi}(p \cdot a)$.*

Proof. We notice that

$$\hat{\pi}(p) \cdot \pi(a) = \hat{\pi}(p) \cdot \hat{\pi}(\mathrm{tp}(a/\mathcal{G})) = \hat{\pi}(p * \mathrm{tp}(a/\mathcal{G})) = \hat{\pi}(p \cdot a),$$

where the second equality is by Lemma 3.5(i). The other computation is similar. □

Lemma 3.8. *Let G be any model of T. Let $h : S_x(G) \to \mathbb{R}$ be a function, $\{[\psi_i]\}_{i \in [n]}$ a partition of $S_x(G)$ with $\psi_i \in \mathcal{L}_x(G)$, $\varepsilon \in \mathbb{R}_{>0}$ and $r_1, \dots, r_n \in \mathbb{R}$ such that $\sup_{q \in S_x(G)} \left| h(q) - \sum_{i=1}^n r_i \mathbf{1}_{[\psi_i]}(q) \right| < \varepsilon$. For $a \in G$, we define the functions $h \cdot a$, $a \cdot h :$ $S_x(G) \to \mathbb{R}$ via $(h \cdot a)(p) = h(p \cdot a)$ and $(a \cdot h)(p) = h(a \cdot p)$. Then*

$$\sup_{q \in S_x(\mathcal{G})} \left| (h \cdot a)(q) - \sum_{i=1}^n r_i \mathbf{1}_{[\psi_i(x \cdot a)]}(q) \right| < \varepsilon,$$

$$\sup_{q \in S_x(\mathcal{G})} \left| (a \cdot h)(q) - \sum_{i=1}^n r_i \mathbf{1}_{[\psi_i(a \cdot x)]}(q) \right| < \varepsilon.$$

In particular, if h is continuous, then $h \cdot a$ and $a \cdot h$ are both continuous maps from $S_x(\mathcal{G})$ to $\mathbb{R}$ (as uniform limits of continuous functions, using in item (2) of Fact 2.14).

Proof. We only prove the lemma for $h \cdot a$ (the case of $a \cdot h$ is similar). Assume the conclusion fails. Then there exists some $q \in S_x(G)$ such that

$$\left| (h \cdot a)(q) - \sum_{i=1}^n r_i \mathbf{1}_{[\psi_i(x \cdot a)]}(q) \right| > \varepsilon.$$

Since $\{[\psi_i(x)]\}_{i \in [n]}$ is a partition, so is $\{[\psi_i(x \cdot a)]\}_{i \in [n]}$. For precisely one $k \in [n]$, we have that $\psi_k(x \cdot a) \in q$ and $\sum_{i=1}^n r_i \mathbf{1}_{[\psi_i(x \cdot a)]}(q) = r_k$. So $\psi_k(x \cdot a^{-1} \cdot a) \in q \cdot a$, and thus $\psi_k(x) \in q \cdot a$. Since $\{[\psi_i(x)]\}_{i \in [n]}$ forms a partition, $\sum_{i=1}^n r_i \mathbf{1}_{[\psi_i(x)]}(q \cdot a) = r_k$. Then $\varepsilon > \left| h(q \cdot a) - \sum_{i=1}^n r_i \mathbf{1}_{[\psi_i(x)]}(q \cdot a) \right| = |(h \cdot a)(q) - r_k| > \varepsilon$ by assumption, a contradiction. □

Remark 3.9. The previous lemma follows also from the more general observation that both the left and right action of $\mathcal{G}$ on $(\mathbb{R}^{S_x(\mathcal{G})}, \| \cdot \|_\infty)$ are by isometries, where $\mathbb{R}^{S_x(\mathcal{G})}$ is the space of all functions from $S_x(\mathcal{G})$ to $\mathbb{R}$ with the uniform norm.

Theorem 3.10. *Suppose $\mu, \nu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$. Then $\pi_*(\mu * \nu) = \pi_*(\mu) \star \pi_*(\nu)$.*

Proof. It suffices to show that for any continuous function $f : \mathcal{G}/\mathcal{G}^{00} \to \mathbb{R}$ we have $\pi_*(\mu * \nu)(f) = \pi_*(\mu) \star \pi_*(\nu)(f)$. Fix a continuous $f : \mathcal{G}/\mathcal{G}^{00} \to \mathbb{R}$. Let $r : S_x(\mathcal{G}) \to S_x(G)$, $p \mapsto p|_G$ be the restriction map. Fix $\varepsilon > 0$. Then $f \circ \hat{\pi}$ is a continuous function from $S_x(\mathcal{G})$ to $\mathbb{R}$ (which factors through $S_x(G)$), so by Fact 2.14(2) there exists a partition $\{[\psi_i(x)]\}_{i \in [n]}$ of $S_x(\mathcal{G})$ with $\psi_i(x) \in \mathcal{L}_x(G)$

and $r_1, \dots, r_n \in \mathbb{R}$ such that

$$\sup_{p \in S_x(\mathcal{G})} \left| (f \circ \hat{\pi})(p) - \sum_{i=1}^{n} r_i \mathbf{1}_{[\psi_i(x)]}(p) \right| < \varepsilon.$$

We now have the following computation for $\pi_*(\mu * \nu)(f)$:

$$\begin{aligned}
\pi_*(\mu * \nu)(f) &= \int_{\mathcal{G}/\mathcal{G}^{00}} f \, d\pi_*(\mu * \nu) = \int_{S_x(\mathcal{G})} (f \circ \hat{\pi}) \, d(\mu * \nu) \\
&\approx_\varepsilon \int_{S_x(\mathcal{G})} \left(\sum_{i=1}^{n} r_i \mathbf{1}_{[\psi_i(x)]} \right) d(\mu * \nu) = \sum_{i=1}^{n} r_i \big((\mu * \nu)(\psi_i(x)) \big) \\
&= \sum_{i=1}^{n} r_i \big((\mu_x \otimes \nu_y)(\psi_i(x \cdot y)) \big) = \sum_{i=1}^{n} r_i \int_{S_y(G)} F_{\mu,G}^{\psi_i'} \, d(\nu_G) \\
&\overset{(*)}{=} \sum_{i=1}^{n} r_i \int_{S_y(\mathcal{G})} (F_{\mu,G}^{\psi_i'} \circ r) \, d\nu = \int_{S_y(\mathcal{G})} \left(\left(\sum_{i=1}^{n} r_i F_{\mu,G}^{\psi_i'} \right) \circ r \right) d\nu.
\end{aligned}$$

The equality $(*)$ is justified by Fact 2.32.

Next we will show that the convolution product $(\pi_*(\mu) \star \pi_*(\nu))(f)$ in $\mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ is close to the final term in the above computation. Define $h : \mathcal{G}/\mathcal{G}^{00} \to \mathbb{R}$ via $h(\boldsymbol{a}) = \int_{\mathcal{G}/\mathcal{G}^{00}} f(x \cdot \boldsymbol{a}) \, d\pi_*(\mu)$. By Lemma 2.12, h is continuous. Fix $p \in S_y(\mathcal{G})$ and let $\boldsymbol{b} := \hat{\pi}(p) \in \mathcal{G}/\mathcal{G}^{00}$ and $b \models r(p) \in \mathcal{G}$. By definition, $\hat{\pi}(p) = \pi(b) = \boldsymbol{b}$. By Lemmas 3.5 and 3.8, we have the following computation:

$$\begin{aligned}
&(h \circ \hat{\pi})(p) = h(\boldsymbol{b}) \\
&= \int_{\mathcal{G}/\mathcal{G}^{00}} f(x \cdot \boldsymbol{b}) \, d\pi_*(\mu) = \int_{q \in S_x(\mathcal{G})} f(\hat{\pi}(q) \cdot \boldsymbol{b}) \, d\mu \\
&= \int_{q \in S_x(\mathcal{G})} f(\hat{\pi}(q) \cdot \pi(b)) \, d\mu = \int_{q \in S_x(\mathcal{G})} f(\hat{\pi}(q \cdot b)) \, d\mu = \int_{S_x(\mathcal{G})} ((f \circ \hat{\pi}) \cdot b) \, d\mu \\
&\approx_\varepsilon \int_{S_x(\mathcal{G})} \sum_{i=1}^{n} r_i \mathbf{1}_{[\psi_i(x \cdot b)]} \, d\mu = \sum_{i=1}^{n} r_i \mu(\psi_i(x \cdot b)) = \left(\left(\sum_{i=1}^{n} r_i F_{\mu,G}^{\psi_i'} \right) \circ r \right)(p).
\end{aligned}$$

Since p was arbitrary in $S_y(\mathcal{G})$, we conclude that

$$\sup_{p \in S_y(\mathcal{G})} \left| (h \circ \hat{\pi})(p) - \left(\left(\sum_{i=1}^{n} r_i F_{\mu,G}^{\psi'} \right) \circ r \right)(p) \right| < \varepsilon.$$

Therefore,

$$\begin{aligned}
(\pi_*(\mu) \star \pi_*(\nu))(f) &= \int_{\mathcal{G}/\mathcal{G}^{00}} h \, d\pi_*(\nu) = \int_{S_y(\mathcal{G})} (h \circ \hat{\pi}) \, d\nu \\
&\approx_\varepsilon \int_{S_y(\mathcal{G})} \left(\left(\sum_{i=1}^{n} r_i F_{\mu,G}^{\psi'} \right) \circ r \right) d\nu \approx_\varepsilon \pi_*(\mu * \nu)(f).
\end{aligned}$$

Since ε was arbitrary, we conclude that $\pi_*(\mu * \nu)(f) = (\pi_*(\mu) \star \pi_*(\nu))(f)$. $\square$

Corollary 3.11. *If $\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ and μ is idempotent, then $\pi_*(\mu)$ is an idempotent measure on $\mathcal{G}/\mathcal{G}^{00}$.*

Proof. By Theorem 3.10 we have $\pi_*(\mu) \star \pi_*(\mu) = \pi_*(\mu * \mu) = \pi_*(\mu)$. □

Corollary 3.12. *Let $\lambda \in \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ and assume that λ is idempotent. Then there exists a measure $\nu \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ such that $\pi_*(\nu) = \lambda$ and ν is idempotent.*

Proof. By Proposition 3.4, the set $A := \{\eta \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G) : \pi_*(\eta) = \lambda\}$ is nonempty. Since π_* is continuous by Fact 3.1(iv), A is a closed subset of $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$. And for any $\eta_1, \eta_2 \in A$ we have $\eta_1 * \eta_2 \in A$, as $\pi_*(\eta_1 * \eta_2) = \pi_*(\eta_1) \star \pi_*(\eta_2) = \lambda \star \lambda = \lambda$ by Theorem 3.10. Hence $(A, *)$ is a compact left-continuous semigroup (using Fact 2.29). By Fact 2.34, $(A, *)$ contains an idempotent. □

4. $\mathcal{G}^{00}$-invariant idempotent measures and type-definable subgroups

In this section we use the properties of the pushforward map established in Section 3 to prove that if μ is idempotent, $\mathcal{G}^{00}$-right-invariant, and automorphism-invariant over a small model, then μ is a translation-invariant measure on its type-definable stabilizer subgroup of $\mathcal{G}$.

Definition 4.1. (1) Let $\mu \in \mathfrak{M}_x(\mathcal{G})$. The *right stabilizer of* μ, denoted as $\mathrm{Stab}(\mu)$, is the subgroup of $\mathcal{G}$ defined by

$$\mathrm{Stab}(\mu) := \bigcap_{\varphi \in \mathcal{L}_x(\mathcal{G})} \{g \in \mathcal{G} : \mu(\varphi(x)) = \mu(\varphi(x \cdot g))\}.$$

(2) Let $\mathcal{H}$ be a subgroup of $\mathcal{G}$ (not necessarily definable). We say that $\mu \in \mathfrak{M}_x(\mathcal{G})$ is *$\mathcal{H}$-right-invariant* (respectively, *$\mathcal{H}$-left-invariant*) if for every formula $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$ and $h \in \mathcal{H}$ we have $\mu(\varphi(x \cdot h)) = \mu(\varphi(x))$ (respectively, $\mu(\varphi(h \cdot x)) = \mu(\varphi(x))$). We say that μ is *$\mathcal{H}$-invariant* if μ is both $\mathcal{H}$-left-invariant and $\mathcal{H}$-right-invariant.

(3) Let $\mathcal{H}$ be a type-definable subgroup of $\mathcal{G}$. We say that $\mathcal{H}$ is *definably amenable* if there exists some $\mu \in \mathfrak{M}_x(\mathcal{G})$ such that $\tilde{\mu}([\mathcal{H}]) = 1$ (where $\tilde{\mu}$ is the unique regular Borel probability measure extending μ) and μ is $\mathcal{H}$-right-invariant. Moreover, in this case we say that $(\mathcal{H}, \mu)$ is an *amenable pair*.

The next proposition shows that if a Keisler measure witnesses the definable amenability of some type-definable subgroup of $\mathcal{G}$, then this subgroup must be its stabilizer:

Proposition 4.2. *Suppose that $\mu \in \mathfrak{M}_x(\mathcal{G})$ and $\mathcal{H}$ is a type-definable subgroup of $\mathcal{G}$. Suppose that $\tilde{\mu}([\mathcal{H}]) = 1$ and $\mathcal{H} \subseteq \mathrm{Stab}(\mu)$. Then $\mathcal{H} = \mathrm{Stab}(\mu)$.*

Proof. Suppose $\mathcal{H} \neq \mathrm{Stab}(\mu)$, and let $g \in \mathrm{Stab}(\mu) \setminus \mathcal{H}$. The subsets $[\mathcal{H}]$ and $[\mathcal{H}] \cdot g$ of $S_x(\mathcal{G})$ are disjoint and $\tilde{\mu}([\mathcal{H}] \cup ([\mathcal{H}] \cdot g)) = 2$, where $\tilde{\mu}$ is the unique regular Borel probability measure extending μ to $S_x(\mathcal{G})$. This is a contradiction. □

Definition 4.3. An idempotent measure $\mu \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ is said to be *pairless* if there does not exist a type-definable subgroup $\mathcal{H}$ of $\mathcal{G}$ such that $(\mathcal{H}, \mu)$ is an amenable pair.

Remark 4.4. By Proposition 4.2, if $\mathrm{Stab}(\mu)$ is type-definable, then μ is pairless if and only if $\mu([\mathrm{Stab}(\mu)]) \neq 1$.

We now give two examples of pairless idempotent measures (in fact, types) in NIP groups (one definable, the other finitely satisfiable). Our third example shows that there can be many measures forming an amenable pair with a given group.

Example 4.5. Let T be the (complete) theory of divisible ordered abelian groups, let $G := (\mathbb{R}, +, <) \models T$, and let $\mathcal{G} \succ G$ be a monster model of T.

(1) Let p_{0^+} be the unique global definable (over $\mathbb{R}$) type extending

$$\{x < a : a > 0, a \in \mathcal{G}\} \cup \{x > a : a \leq 0, a \in \mathbb{R}\}.$$

Then $\delta_{p_{0^+}} \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ is idempotent and pairless.

(2) Let $p_{\mathbb{R}^+}$ be the unique global type finitely satisfiable in $\mathbb{R}$ and extending

$$\{x > a : a \in \mathbb{R}\}.$$

Then $\delta_{p_{\mathbb{R}^+}} \in \mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, G)$ is idempotent and pairless.

(3) Let $p_{+\infty}$ and $p_{-\infty}$ be the unique global heirs (over $\mathbb{R}$) extending the types

$$\Theta_+(x) := \{x > a : a \in \mathbb{R}\} \quad \text{and} \quad \Theta_-(x) := \{x < a : a \in \mathbb{R}\},$$

respectively. Then $(\mathcal{G}, \mu_r)$ is an amenable pair for any $r \in [0, 1]$, where

$$\mu_r = r\delta_{p_{-\infty}} + (1-r)\delta_{p_{+\infty}}.$$

Proof. (1) Note that $\mathrm{Stab}(\delta_{p_{0^+}}) = \{0\}$ and $\delta_{p_{0^+}}(\{0\}) = 0$, so $\delta_{p_0^+}$ is pairless by Proposition 4.2. We now check that $\delta_{p_{0^+}}$ is idempotent. Fix some $a \in \mathcal{G}$, some small $G' \prec \mathcal{G}$ containing a and $\mathbb{R}$, and a realization $c \models p_{0^+}|_{G'}$ in $\mathcal{G}$. Note that

$$(p_{0^+} * p_{0^+})(x < a) = (p_{0^+} \otimes p_{0^+})(x + y < a) = p_{0^+}(x < a - c).$$

We now have two cases:

(a) If $a > 0$, then $a - c > 0$ and so $(p_{0^+} * p_{0^+})(x < a) = 1$.

(b) If $a \leq 0$, then $a - c < 0$ and so $(p_{0^+} * p_{0^+})(x < a) = 0$.

Hence, using quantifier-elimination, $p_{0^+} * p_{0^+} = p_{0^+}$, and so $\delta_{p_{0^+}} * \delta_{p_{0^+}} = \delta_{p_{0^+}}$.

(2) The measure $\delta_{p_{\mathbb{R}^+}}$ is idempotent by a computation analogous to the one in (1). We have

$$\mathrm{Stab}(\delta_{p_{\mathbb{R}^+}}) = \{a \in \mathcal{G} : -n < a < n \text{ for some n} \in \mathbb{N}\}.$$

We note that $\mathrm{Stab}(\delta_{p_{\mathbb{R}^+}})$ is a $\bigvee$-definable subset of $\mathcal{G}$, but is not definable, so it is not type-definable. Now suppose that there exists a type-definable subgroup $\mathcal{H}$

of $\mathcal{G}$ such that $(\mathcal{H}, \delta_{p_{\mathbb{R}+}})$ is an amenable pair. Then, by definition, $\mathcal{H} \subseteq \mathrm{Stab}(\delta_{p_{\mathbb{R}+}})$ and $\delta_{p_{\mathbb{R}+}}([\mathcal{H}]) = 1$. By Proposition 4.2, we conclude that $\mathcal{H} = \mathrm{Stab}(\delta_{p_{\mathbb{R}+}})$. Hence $\mathrm{Stab}(\delta_{p_{\mathbb{R}+}})$ is type-definable, a contradiction. Alternatively, we get a contradiction by regularity of the measure:

$$\delta_{p_{\mathbb{R}+}}([\mathrm{Stab}(\delta_{p_{\mathbb{R}+}})]) = \sup\{\delta_{p_{\mathbb{R}+}}([-n < x < n]) : n \in \mathbb{N}\} = 0.$$

(3) Note that $p_{+\infty}$ and $p_{-\infty}$ are (left- and right-) $\mathcal{G}$-invariant. Hence

$$\mu_r := r\delta_{p_{-\infty}} + (1-r)\delta_{p_{+\infty}} \in \mathfrak{M}_x(\mathcal{G})$$

is $\mathcal{G}$-invariant for any $r \in [0, 1]$. Since μ_r is $\mathcal{G}$-invariant, $(\mathcal{G}, \mu_r)$ is an amenable pairing for every $r \in [0, 1]$. □

In the rest of this section we show that in an NIP group $\mathcal{G}$, for any $\mathcal{G}^{00}$-invariant idempotent $\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$, $\mathrm{Stab}(\mu)$ is type-definable and $(\mathrm{Stab}(\mu), \mu)$ is an amenable pair.

Definition 4.6. Assume that $\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ is idempotent. By Corollary 3.11, the measure $\pi_*(\mu) \in \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ is idempotent, and by Fact 2.8, $\mathrm{supp}(\pi_*(\mu))$ is a closed subgroup of $\mathcal{G}/\mathcal{G}^{00}$ and $\pi_*(\mu) \restriction_{\mathrm{supp}(\pi_*(\mu))}$ is the normalized Haar measure on this closed subgroup. Then $\pi^{-1}\big(\mathrm{supp}(\pi_*(\mu))\big)$ is a type-definable subgroup of $\mathcal{G}$. We let $H_{\mathcal{L}}(\mu) := \pi^{-1}\big(\mathrm{supp}(\pi_*(\mu))\big)$.

Proposition 4.7. *Suppose $\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ is idempotent and $\mathcal{G}^{00}$-right-invariant.*

(i) *If $p \in \sup(\mu)$, then $\hat{\pi}(p) \in \mathrm{supp}(\pi_*(\mu))$* (see Fact 3.1 for the definition of $\hat{\pi}$).

(ii) *If $p \in \sup(\mu)$, then $p \in [H_{\mathcal{L}}(\mu)]$.*

(iii) *$\mu([H_{\mathcal{L}}(\mu)]) = 1$.*

(iv) *If $b \in \mathrm{Stab}(\mu)$, then $\pi(b) \in \mathrm{Stab}(\pi_*(\mu))$.*

Proof. (i) Let U be an open subset of $\mathcal{G}/\mathcal{G}^{00}$ containing $\hat{\pi}(p)$. Then $\pi^{-1}(U)$ is $\bigvee$-definable, so $\pi^{-1}(U) = \bigvee_{i \in I} \psi_i(x)$ for some $\psi_i \in \mathcal{L}_x(G)$. Hence there exists some $i \in I$ so that $\psi_i(x) \in p$. Since $p \in \sup(\mu)$, we have that $\mu(\psi_i(x)) > 0$. Then

$$\pi_*(\mu)(U) = \tilde{\mu}([\hat{\pi}^{-1}(U)]) \geq \mu(\psi_i(x)) > 0,$$

where $\tilde{\mu}$ is the unique regular Borel probability measures extending μ. Therefore $\hat{\pi}(p) \in \mathrm{supp}(\pi_*(\mu))$.

(ii) This is obvious by (i).

(iii) Assume not. Then $\mu(S_x(\mathcal{G})\backslash[H_{\mathcal{L}}(\mu)]) > 0$. This set is open and so by regularity there exists some $[\psi(x)] \subset S_x(\mathcal{G}) \setminus [\mathcal{H}_{\mathcal{L}}(\mu)]$ such that $\mu(\psi(x)) > 0$. Then there exists some $p \in \sup(\mu)$ so that $\psi(x) \in p$. This contradicts (ii).

(iv) By Theorem 3.10,

$$\pi_*(\mu) \cdot \pi(b) = \pi_*(\mu) \star \delta_{\pi(b)} = \pi_*(\mu * \delta_b) = \pi_*(\mu).$$ □

Lemma 4.8. *Assume that $f : S_x(\mathcal{G}) \to \mathbb{R}$ is Borel and factors through $\hat{\pi} : S_x(\mathcal{G}) \to \mathcal{G}/\mathcal{G}^{00}$, and let $f_\star : \mathcal{G}/\mathcal{G}^{00} \to \mathbb{R}$ be the factor map. Then $f_\star$ is Borel.*

Proof. The map $\hat{\pi} : S_x(\mathcal{G}) \to \mathcal{G}/\mathcal{G}^{00}$ is a continuous surjective map between compact Hausdorff spaces. If the map $f = f_\star \circ \hat{\pi}$ is Borel, then $f_\star$ is Borel by [Holický and Spurný 2003, Theorem 10] (see [Conant et al. 2021, Theorem 2.1] for an explanation). □

Lemma 4.9. *Assume that $\mu \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ is idempotent and $\mathcal{G}^{00}$-right-invariant. Suppose that $p \in \sup(\mu|_G)$ and $a \models p$ in $\mathcal{G}$. Then $\mu(\varphi(x)) = \mu(\varphi(x \cdot a))$ for any $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$.*

Proof. Fix $p \in \sup(\mu|_G)$, $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$ and $a \in \mathcal{G}$ such that $a \models p$. Fix a small model $G' \prec \mathcal{G}$ such that G' contains G, a, and all of the parameters of φ. Let $r : S_y(\mathcal{G}) \to S_y(G')$, $q \mapsto q|_{G'}$ be the restriction map. Since μ is idempotent,

$$\mu(\varphi(x \cdot a)) = \mu * \mu(\varphi(x \cdot a)) = \int_{S_y(G')} F^{\varphi'_a}_{\mu,G'}\, d\mu_{G'} = \int_{S_y(\mathcal{G})} (F^{\varphi'_a}_{\mu,G'} \circ r)\, d\mu,$$

where $\varphi_a(x) := \varphi(x \cdot a)$, so $\varphi'_a(x, y) = \varphi(x \cdot y \cdot a)$ and $F^{\varphi'_a}_{\mu,G'}(q) = \mu(\varphi(x \cdot c \cdot a))$ for some/any $c \models q$ (see Definition 2.26). Let $f := F^{\varphi'_a}_{\mu,G'} \circ r$ and $h := F^{\varphi'}_{\mu,G'} \circ r$.

Claim 1: Both f and h factor through $\hat{\pi} : S_y(\mathcal{G}) \to \mathcal{G}/\mathcal{G}^{00}$.

Proof. The proofs are essentially the same, so we only show that f factors through $\hat{\pi}$. Fixing $q_1, q_2 \in S_y(\mathcal{G})$ with $\hat{\pi}(q_1) = \hat{\pi}(q_2)$, we want to show that then $f(q_1) = f(q_2)$. Let $b_1, b_2 \in \mathcal{G}$ be such that $b_1 \models r(q_1)$ and $b_2 \models r(q_2)$. Then $\pi(b_1) = \pi(b_2)$. Since $\mathcal{G}^{00}$ is a normal subgroup of $\mathcal{G}$, we then have $b_1 = d \cdot b_2$ for some $d \in \mathcal{G}^{00}$. Hence

$$f(q_1) = (F^{\varphi'_a}_{\mu,G'} \circ r)(q_1) = \mu(\varphi(x \cdot b_1 \cdot a)) = \mu(\varphi(x \cdot d \cdot b_2 \cdot a)).$$

And since μ is $\mathcal{G}^{00}$-right-invariant, we have $\mu(\psi(x \cdot d)) = \mu(\psi(x))$ for $\psi(x) := \varphi(x \cdot b_2 \cdot a)$, that is,

$$\mu(\varphi(x \cdot d \cdot b_2 \cdot a)) = \mu(\varphi(x \cdot b_2 \cdot a)) = (F^{\varphi'_a}_{\mu,G'} \circ r)(q_2) = f(q_2). \qquad \square$$

We let $f_\star$ and $h_\star$ be the associated factor maps from $\mathcal{G}/\mathcal{G}^{00}$ to $\mathbb{R}$.

Claim 2: We have $h_\star \cdot \pi(a) = f_\star$, where $h_\star \cdot \pi(a) : \mathcal{G}/\mathcal{G}^{00} \to \mathbb{R}$ is the function defined by $(h_\star \cdot \pi(a))(\boldsymbol{b}) := h_\star(\boldsymbol{b} \cdot \pi(a))$ for any $\boldsymbol{b} \in \mathcal{G}/\mathcal{G}^{00}$.

Proof. Fix $\boldsymbol{b} \in \mathcal{G}/\mathcal{G}^{00}$ and $b \in \mathcal{G}$ such that $\pi(b) = \boldsymbol{b}$. Then

$$\begin{aligned}
&(h_\star \cdot \pi(a))(\boldsymbol{b}) \\
&\quad = (h_\star)(\boldsymbol{b} \cdot \pi(a)) = (h_\star)(\pi(b \cdot a)) = (F^{\varphi'}_{\mu,G'} \circ r)(\mathrm{tp}(b \cdot a/\mathcal{G})) \\
&\quad = F^{\varphi'}_{\mu,G'}(\mathrm{tp}(b \cdot a/G')) = \mu(\varphi(x \cdot b \cdot a)) = F^{\varphi'_a}_{\mu,G'}(\mathrm{tp}(b/G')) = f_\star(\boldsymbol{b}). \qquad \square
\end{aligned}$$

Claim 3: $$\mu(\varphi(x \cdot a)) = \mu(\varphi(x)).$$

Proof. The maps $f_\star, h_\star : \mathcal{G}/\mathcal{G}^{00} \to \mathbb{R}$ are Borel by Lemma 4.8. By assumption $a \models p$ with $p \in \sup(\mu|_G)$. Then there exists $\hat{p} \in \sup(\mu)$ such that $\hat{p}|_G = p$ (see, e.g., [Chernikov and Gannon 2022, Proposition 2.8]). By Proposition 4.7(i) we then have $\pi(a) = \hat{\pi}(\hat{p}) \in \operatorname{supp}(\pi_*(\mu))$. The measure $\pi_*(\mu)$ is idempotent by Corollary 3.11. Applying Fact 2.11 (to the compact group $\mathcal{G}/\mathcal{G}^{00}$ and its closed subgroup $\operatorname{supp}(\pi_*(\mu)) \ni \pi(a)$) we get

$$\int_{\mathcal{G}/\mathcal{G}^{00}} (h_\star \cdot \pi(a))\, d\pi_*(\mu) = \int_{\mathcal{G}/\mathcal{G}^{00}} h_\star\, d\pi_*(\mu).$$

Using this and Claim 2 we have the following computation:

$$\begin{aligned} \mu(\varphi(x \cdot a)) &= (\mu * \mu)(\varphi(x \cdot a)) \\ &= \int_{S_y(\mathcal{G})} f\, d\mu = \int_{\mathcal{G}/\mathcal{G}^{00}} f_\star\, d\pi_*(\mu) \\ &= \int_{\mathcal{G}/\mathcal{G}^{00}} (h_\star \cdot \pi(a))\, d\pi_*(\mu) = \int_{\mathcal{G}/\mathcal{G}^{00}} h_\star\, d\pi_*(\mu) \\ &= \int_{S_y(\mathcal{G})} h\, d\mu = \int_{S_y(G')} F^{\varphi'}_{\mu,G'}\, d\mu_{G'} = (\mu * \mu)(\varphi(x)) = \mu(\varphi(x)). \quad \square \end{aligned}$$

This concludes the proof of Lemma 4.9. □

Lemma 4.10. *Suppose that $\boldsymbol{g} \in \operatorname{supp}(\pi_*(\mu))$. Then there exists some $p \in \sup(\mu|_G)$ such that for any $b \models p$ we have $\pi(b) = \boldsymbol{g}$.*

Proof. We use the fact that $\pi_* : \mathcal{M}(S_x(\mathcal{G})) \to \mathcal{M}(\mathcal{G}/\mathcal{G}^{00})$ is a pushforward map. Let $\tilde{\mu}$ be the unique extension of μ to a regular Borel probability measure on $S_x(\mathcal{G})$. Let $\boldsymbol{g} \in \operatorname{supp}(\pi_*(\mu))$ and let $U \subseteq \mathcal{G}/\mathcal{G}^{00}$ be an open set containing $\boldsymbol{g}$. Because $\boldsymbol{g} \in \operatorname{supp}(\pi_*(\mu))$, we have that $0 < \pi^*(\mu)(U) = \tilde{\mu}([\pi^{-1}(U)])$. Then there exists some $p_U \in \operatorname{supp}(\tilde{\mu})$ such that $p_U \in [\pi^{-1}(U)]$. The collection of open sets in $\mathcal{G}/\mathcal{G}^{00}$ containing $\boldsymbol{g}$ forms a directed family under reverse inclusion, and we can consider the net $(p_U)_{\boldsymbol{g} \in U}$. Since $\operatorname{supp}(\tilde{\mu})$ is closed and hence compact, there exists a convergent subnet $(q_i)_{i \in I}$ with a limit in $\operatorname{supp}(\tilde{\mu})$. Let $q := \lim_{i \in I} q_i$. By continuity of $\hat{\pi} : S_x(\mathcal{G}) \to \mathcal{G}/\mathcal{G}^{00}$, we have that $\hat{\pi}(q) = \boldsymbol{g}$. Since $\sup(\mu) = \operatorname{supp}(\tilde{\mu})$ we conclude that $q \in \sup(\mu)$. By definition of $\hat{\pi}$ we have $\hat{\pi}(q) = \pi(b)$ for any $b \models q|_G$, so the lemma holds with $p := q|_G$. □

Theorem 4.11. *Suppose that $\mu \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ is idempotent and $\mathcal{G}^{00}$-right-invariant. Then*

(1) $\operatorname{Stab}(\mu) = H_{\mathcal{L}}(\mu)$ (see Definition 4.6);

(2) $\operatorname{Stab}(\mu)$ *is a type-definable subgroup of* $\mathcal{G}$;

(3) $(\operatorname{Stab}(\mu), \mu)$ *is an amenable pair.*

Proof. (1) As $H_{\mathcal{L}}(\mu)$ is a type-definable subgroup of $\mathcal{G}$, by Proposition 4.2 it suffices to show that μ is $H_{\mathcal{L}}(\mu)$-right-invariant and $\mu([H_{\mathcal{L}}(\mu)])=1$. By Proposition 4.7(iii), we have $\mu([H_{\mathcal{L}}(\mu)]) = 1$, so it remains to show that $H_{\mathcal{L}}(\mu) \subseteq \mathrm{Stab}(\mu)$. Fix $a \in H_{\mathcal{L}}(\mu)$. Then $\boldsymbol{g} := \pi(a) \in \mathrm{supp}(\pi_*(\mu))$ by Proposition 4.7(i). By Lemma 4.10, there exists some $p \in \sup(\mu|_G)$ and $b \models p$ such that $\pi(b) = \boldsymbol{g}$. In particular, $a \cdot \mathcal{G}^{00} = b \cdot \mathcal{G}^{00}$, so $a = c \cdot b$ for some $c \in \mathcal{G}^{00}$. Now we have

$$\mu(\varphi(x \cdot a)) = \mu(\varphi(x \cdot c \cdot b)) = \mu(\varphi(x \cdot b)) = \mu(\varphi(x)).$$

The second equality follows from the fact that μ is $\mathcal{G}^{00}$-right-invariant and the fourth equality follows from Lemma 4.9.

(2) This follows from the fact that $\mathrm{Stab}(\mu) = \mathcal{H}_{\mathcal{L}}(\mu)$ and $\mathcal{H}_{\mathcal{L}}(\mu)$ is type-definable.

(3) This follows since $\mu([\mathrm{Stab}(\mu)]) = \mu([\mathcal{H}_{\mathcal{L}}(\mu)]) = 1$. □

5. The structure of convolution semigroups

By Fact 2.29, if T is an NIP theory expanding a group, then both $(\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G), *)$ and $(\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G), *)$ are left-continuous compact Hausdorff semigroups (and hence satisfy the assumption of Fact 2.34). In this section we describe some properties of the minimal left ideals and ideal groups which arise in this setting. Unlike the better studied case of the semigroup $(S_x^{\mathrm{fs}}(\mathcal{G}, G), *)$, we demonstrate that the ideal subgroups of any minimal left ideal (in either $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ or $\mathfrak{M}^{\mathrm{inv}}(\mathcal{G}, G)$) are always trivial, i.e., isomorphic to the group with a single element. The following theorem summarizes the properties that we will prove in this section.

Theorem 5.1. *Assume that $\mathcal{G}$ is NIP, and let I be a minimal left ideal of $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$* (which exists by Fact 2.34). *Then we have the following*:

(1) *I is a closed convex subset of $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$* (Proposition 5.3).

(2) *For any $\mu \in I$, $\pi_*(\mu) = h$, where h is the normalized Haar measure on $\mathcal{G}/\mathcal{G}^{00}$* (Proposition 5.5).

(3) *If $\mathcal{G}/\mathcal{G}^{00}$ is nontrivial, then I does not contain any types* (Proposition 5.7).

(4) *For any idempotent $u \in I$, we have $u * I \cong (e, \cdot)$. In other words, the ideal group is always trivial* (Proposition 5.10).

(5) *Every element of I is an idempotent* (Proposition 5.11).

(6) *If $\mu, \nu \in I$ then $\mu * \nu = \mu$* (Proposition 5.11).

(7) *For any $\mu \in I$, $I = \{\nu \in \mathfrak{M}_x^{\dagger}(\mathcal{G}, G) : \nu * \mu = \nu\}$* (Corollary 5.12).

(8) *For any definable measure $\nu \in \mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$ there exists a measure $\mu \in I$ such that $\nu * \mu = \mu$. In particular, for any $g \in G$ there exists a measure $\mu \in I$ such that $\delta_g * \mu = g \cdot \mu = \mu$* (Proposition 5.13).

(9) *Assume that $\mathcal{G}$ is definably amenable.*

(a) *If* $\dagger = \mathrm{fs}$, *then* $I = \{\nu\}$, *where* $\nu \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ *is a G-left-invariant measure* (Proposition 5.16).

(b) *If* $\dagger = \mathrm{inv}$, *then*

$$I = \{\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G) : \mu \textit{ is } \mathcal{G}\textit{-right-invariant}\}.$$

Moreover, I is a two-sided ideal, and is the unique minimal left ideal (Proposition 5.18). *The set* $\mathrm{ex}(I)$ *of extreme points of I is closed and equal to $\{\mu_p : p \in S_x^{\mathrm{inv}}(\mathcal{G}, G)$ is right f-generic$\}$, and I is a Bauer simplex* (Corollary 5.21).

(10) *If $\mathcal{G}$ is fsg and $\mu \in \mathfrak{M}_x(\mathcal{G})$ is the unique $\mathcal{G}$-left-invariant measure, then $I = \{\mu\}$ is the unique minimal left (in fact, two-sided) ideal in both $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ and $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$* (Corollary 5.24).

(11) *If $\mathcal{G}$ is not definably amenable, then the closed convex set I has infinitely many extreme points* (Remark 5.26).

We remark that (5) and (11) of Theorem 5.1 guarantee the existence of many idempotent measures in nondefinably amenable NIP groups. All previous "constructions" of idempotent measures either explicitly or implicitly use definable amenability or amenability of closed subgroups of $\mathcal{G}/\mathcal{G}^{00}$. A priori, the idempotent measures we find here have no connection to type-definable subgroups.

5A. *General structure.* Our first goal is to show that any minimal left ideal of $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$ is convex. We begin by showing that convolution is affine in both arguments and therefore preserves convexity on both sides.

Lemma 5.2. *Assume $\mu, \lambda_1, \lambda_2 \in \mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$ and $r, s \in \mathbb{R}_{>0}$ with $r + s = 1$. We have:*

(1) $(r\lambda_1 + s\lambda_2) * \mu = r(\lambda_1 * \mu) + s(\lambda_2 * \mu)$.

(2) $\mu * (r\lambda_1 + s\lambda_2) = r(\mu * \lambda_1) + s(\mu * \lambda_2)$.

(3) *If $A \subseteq \mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$ is convex, then both $\mu * A$ and $A * \mu$ are convex.*

Proof. Parts (1) and (2) were stated in [Chernikov and Gannon 2022, Proposition 3.14(4)], but no proof was provided there, so we take the opportunity to provide it here.

(1) Fix a formula $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$ and let G' be a small model containing G and the parameters of φ. Then

$$((r\lambda_1 + s\lambda_2) * \mu)(\varphi(x)) = \int_{S_y(G')} F^{\varphi'}_{r\lambda_1 + s\lambda_2} \, d\mu_{G'} = \int_{S_y(G')} (rF^{\varphi'}_{\lambda_1} + sF^{\varphi'}_{\lambda_2}) \, d\mu_{G'}$$

$$\begin{aligned}&= r\int_{S_y(G')} F^{\varphi'}_{\lambda_1}\, d\mu_{G'} + s\int_{S_y(G')} F^{\varphi'}_{\lambda_2}\, d\mu_{G'}\\ &= r(\lambda_1 * \mu)(\varphi(x)) + s(\lambda_2 * \mu)(\varphi(x))\\ &= (r(\lambda_1 * \mu) + s(\lambda_2 * \mu))(\varphi(x)).\end{aligned}$$

(2) Fix a formula $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$ and a small model G' containing G and the parameters of φ. Then the map $F^{\varphi'}_{\mu} : S_y(G') \to [0, 1]$ is a bounded Borel function, so for any $\varepsilon > 0$ there exist Borel subsets $B_1, \dots, B_n$ of $S_y(G')$ and real numbers $k_1, \dots, k_n$ such that

$$\sup_{q\in S_y(G)} \left| F^{\varphi'}_{\mu}(q) - \sum_{i=1}^{n} k_i \mathbf{1}_{B_i}(q)\right| < \varepsilon.$$

Now we compute the product:

$$\begin{aligned}&(\mu * (r\lambda_1 + s\lambda_2))(\varphi(x))\\ &\qquad = \int_{S_y(G')} F^{\varphi'}_{\mu}\, d(r\lambda_1 + s\lambda_2)\\ &\qquad \approx_\varepsilon \int_{S_y(G')} \left(\sum_{i=1}^{n} k_i \mathbf{1}_{B_i}\right) d(r\lambda_1 + s\lambda_2) = r\sum_{i=1}^{n} k_i\lambda_1(B_i) + s\sum_{i=1}^{n} k_i\lambda_2(B_i)\\ &\qquad = r\int_{S_y(G')} \left(\sum_{i=1}^{n} k_i \mathbf{1}_{B_i}\right) d\lambda_1 + s\int_{S_y(G')} \left(\sum_{i=1}^{n} k_i \mathbf{1}_{B_i}\right) d\lambda_2\\ &\qquad \approx_\varepsilon r\int_{S_y(G')} F^{\varphi'}_{\mu}\, d\lambda_1 + s\int_{S_y(G')} F^{\varphi'}_{\mu}\, d\lambda_2 = (r(\mu * \lambda_1) + s(\mu * \lambda_2))(\varphi(x)).\end{aligned}$$

(3) We first prove that $A * \mu$ is convex. Letting $\nu_1, \nu_2 \in A * \mu$ and $r, s \in \mathbb{R}_{>0}$ with $r + s = 1$ be given, we need to show that $r\nu_1 + s\nu_2 \in A * \mu$. By assumption there exist some $\lambda_1, \lambda_2 \in A$ such that $\lambda_i * \mu = \nu_i$ for $i \in \{1, 2\}$. Since A is convex, we have that $r\lambda_1 + s\lambda_2 \in A$. It follows by (1) that $r\nu_1 + s\nu_2 = (r\lambda_1 + s\lambda_2) * \mu \in A * \mu$.

Now we prove that $\mu * A$ is convex. Similarly, let $\nu_1, \nu_2 \in \mu * A$ and $r, s \in \mathbb{R}_{>0}$ with $r + s = 1$ be given, and let $\lambda_1, \lambda_2 \in A$ be such that $\mu * \lambda_i = \nu_i$. Consider the measure $r\lambda_1 + s\lambda_2 \in A$. It follows by (2) that $r\nu_1 + s\nu_2 = \mu * (r\lambda_1 + s\lambda_2) \in \mu * A$. □

Proposition 5.3. *If I is a minimal left ideal in $\mathfrak{M}^{\dagger}_x(\mathcal{G}, G)$, then I is closed and convex.*

Proof. Any minimal left ideal is closed by Fact 2.34. Choose $\mu \in I$. By Fact 2.34(5), we have $\mathfrak{M}^{\dagger}_x(\mathcal{G}, G) * \mu = I$. By Lemma 5.2 and the convexity of $\mathfrak{M}^{\dagger}_x(\mathcal{G}, G)$, I is convex. □

We now consider the interaction between the pushforward map to $\mathcal{G}/\mathcal{G}^{00}$ and the minimal left ideal. The following lemma is standard.

Lemma 5.4. *Let S be a semigroup, L a minimal left ideal of S, and H a two-sided ideal in S. Then $L \subseteq H$.*

Proof. Note that $L' := L \cap H$ is nonempty (for $l \in L$ and $h \in H$, $h \cdot l \in L \cap H$) and is a left ideal (as an intersection of two left ideals). As $L' \subseteq L$, by minimality $L = L' \subseteq H$. □

Proposition 5.5. *Let I be a minimal left ideal in $\mathfrak{M}_x^\dagger(\mathcal{G}, G)$. Then for every $\nu \in I$ we have $\pi_*^\dagger(\nu) = h$, where h is the normalized Haar measure on $\mathcal{G}/\mathcal{G}^{00}$.*

Proof. Since $\pi_*^\dagger$ is surjective (Proposition 3.4) and continuous (Remark 3.3), the set $A := (\pi_*^\dagger)^{-1}(\{h\})$ is a nonempty closed subset of $\mathfrak{M}_x^\dagger(\mathcal{G}, G)$. Moreover, A is a two-sided ideal: since $\pi_*^\dagger$ is a homomorphism (Theorem 3.10) and h is both left- and right-invariant, for any $\mu \in A$ and $\nu \in \mathfrak{M}_x^\dagger(\mathcal{G}, G)$, we have

$$\pi_*^\dagger(\nu * \mu) = \pi_*^\dagger(\nu) \star \pi_*^\dagger(\mu) = \pi_*^\dagger(\nu) \star h = h,$$

and a similar computation also shows that A is a right ideal. By Lemma 5.4 we have $I \subseteq A$, which completes the proof. □

Definition 5.6. Let $\mu \in \mathfrak{M}_x(\mathcal{G})$. We say μ is *strongly continuous* if for every $\varepsilon > 0$, there exists a finite partition $\{[\psi(x)]\}_{i<n}$ of $S_x(\mathcal{G})$ with $\psi_i \in \mathcal{L}_x(\mathcal{G})$ such that $\mu(\psi(x)) < \varepsilon$ for all $i < n$.

Proposition 5.7. *Let I be a minimal left ideal in $\mathfrak{M}_x^\dagger(\mathcal{G}, G)$.*

(1) *If $\mathcal{G}/\mathcal{G}^{00}$ is nontrivial, then I does not contain any types.*

(2) *If $\mathcal{G}/\mathcal{G}^{00}$ is infinite, then every measure in I is strongly continuous.*

Proof. (1) By Lemma 3.5(2) we have $\pi_*^\dagger(\delta_p) = \delta_{\hat{\pi}(p)}$, which does not equal the normalized Haar measure on $\mathcal{G}/\mathcal{G}^{00}$ when it is nontrivial. This contradicts Proposition 5.5.

(2) If $\mathcal{G}/\mathcal{G}^{00}$ is infinite then the normalized Haar measure h on $\mathcal{G}/\mathcal{G}^{00}$ is zero on every point. Suppose that $\nu \in \mathfrak{M}_x^\dagger(\mathcal{G}, G)$ is not strongly continuous. By compactness and [Bhaskara Rao and Bhaskara Rao 1983, Theorem 5.2.7], ν can be written as

$$\nu = r_0\mu_0 + \sum_{i\in\omega} r_i \delta_{p_i},$$

where $\mu_0 \in \mathfrak{M}_x^\dagger(\mathcal{G}, G)$ is strongly continuous, $r_i \in [0, 1]$ and $p_i \in S_x^\dagger(\mathcal{G}, G)$ for each $i \in \omega$, and $\sum_{i\in\omega} r_i = 1$. We then must have $r_{i^*} > 0$ for some $i^* \in \omega\backslash\{0\}$. Since the pushforward map is affine (Remark 3.3), we have

$$\pi_*(\nu) = r_0\pi_*(\mu_0) + \sum_{i\in\omega} r_i \delta_{\hat{\pi}(p_i)}.$$

Hence $\pi_*(\nu)(\{\hat{\pi}(p_{i^*})\}) = r_{i^*} > 0$, so $\pi_*(\nu) \neq h$, contradicting Proposition 5.5. □

We now show that the ideal subgroup of any minimal left ideal is trivial. A related result appears in [Cohen and Collins 1959, Theorem 3], but we are working in a semigroup which is only left-continuous. Our proof is a generalization of the proof that there do not exist any nontrivial convex compact groups and follows [Abodayeh and Murphy 1997, Lemmas 3.1 and 3.2]. In particular, compactness is used only to get an extreme point in *some* ideal subgroup. Some elementary algebra is then used to show that the only possible ideal subgroups are isomorphic to a single point.

Lemma 5.8. *If I is a minimal left ideal in $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$, then* $\mathrm{ex}(I) \neq \varnothing$.

Proof. By Proposition 5.3, I is a compact convex set. By the Krein–Milman theorem, I contains an extreme point. □

Lemma 5.9. *If I is a minimal left ideal in $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$, then there exists an idempotent μ in I such that $\mu \in \mathrm{ex}(\mu * I)$.*

Proof. By Lemma 5.8, there exists a measure $\nu \in I$ which is extreme in I. By Fact 2.34(4), there exists an idempotent μ in I such that $\nu \in \mu * I$. Towards a contradiction, suppose that $\mu \notin \mathrm{ex}(\mu * I)$. Then there exist distinct $\eta_1, \eta_2 \in \mu * I$ and $r \in (0, 1)$ such that $r\eta_1 + (1-r)\eta_2 = \mu$. As μ is the identity of the group $\mu * I$ by Fact 2.34(3), we get

$$\nu = \nu * \mu = r(\nu * \eta_1) + (1-r)(\nu * \eta_2).$$

Since $\nu \in \mathrm{ex}(I)$ and $\nu * \eta_i \in I$ as I is a left ideal, it follows that $\nu = \nu * \eta_1 = \nu * \eta_2$. Since $\nu, \eta_1, \eta_2 \in \mu * I$ and $\mu * I$ is a group, this implies $\eta_1 = \eta_2$, contradicting the assumption. Hence $\mu \in \mathrm{ex}(\mu * I)$. □

Proposition 5.10. *The ideal subgroup of $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$ is trivial.*

Proof. Let I be a minimal left ideal of $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$. By Lemma 5.9, there exists an idempotent $\mu \in I$ such that μ is extreme in $\mu * I$. Let $\eta_1, \eta_2 \in \mu * I$. We will show that $\eta_1 = \eta_2$. By Lemma 5.2 and Proposition 5.3, $\mu * I$ is convex. Hence $\alpha := \frac{1}{2}(\eta_1 + \eta_2) \in \mu * I$. Since $\mu * I$ is a group with identity μ, $\mu * I$ contains α^{-1} (i.e., $\alpha^{-1} * \alpha = \alpha * \alpha^{-1} = \mu$). Then

$$\mu = \alpha^{-1} * \alpha = \alpha^{-1} * \left(\tfrac{1}{2}\eta_1 + \tfrac{1}{2}\eta_2\right) = \tfrac{1}{2}(\alpha^{-1} * \eta_1) + \tfrac{1}{2}(\alpha^{-1} * \eta_2).$$

Since μ is extreme in $\mu * I$ and $\alpha^{-1} * \eta_i \in \mu * I$, we get $\mu = \alpha^{-1} * \eta_1 = \alpha^{-1} * \eta_2$ and hence $\eta_1 = \eta_2$. □

We have shown that any ideal subgroup of $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$ is trivial. Since the minimal left ideals can be partitioned into their ideal subgroups, it follows that the convolution operation is trivial when restricted to a minimal left ideal.

Proposition 5.11. *Let I be a minimal left ideal in $\mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$. Then every element of I is an idempotent. Moreover, for any elements $\mu, \nu \in I$, we have that $\mu * \nu = \mu$.*

Proof. By Fact 2.34(4) and Proposition 5.10,

$$I = \bigsqcup_{\mu \in \mathrm{id}(I)} \mu * I = \bigsqcup_{\mu \in \mathrm{id}(I)} \{\mu\} = \mathrm{id}(I).$$

The "moreover" part also follows from the observation that $\mu * I = \{\mu\}$. □

Corollary 5.12. *Let I be a minimal left ideal of $\mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$ and assume that $\mu \in I$. Then $I = \{\nu \in \mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G) : \nu * \mu = \nu\}$.*

Proof. By Proposition 5.11 we have $I \subseteq \{\nu \in \mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G) : \nu * \mu = \nu\}$. And since I is a left ideal and $\mu \in I$, we have $\{\nu \in \mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G) : \nu * \mu = \nu\} \subseteq I$. □

We also observe that the action of the underlying group G on the minimal left ideal is far from being a free action (this is of course trivial in the definably amenable case, but is meaningful when $\mathcal{G}$ is not definably amenable).

Proposition 5.13. *Let I be a minimal left ideal of $\mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$. For any definable measure $\nu \in \mathfrak{M}^{\dagger}(\mathcal{G}, G)$ there exists a measure $\mu \in I$ such that $\nu * \mu = \mu$. In particular, for every element $g \in G$, there exists a measure $\mu \in I$ such that $\delta_g * \mu = \mu$.*

Proof. Consider the map $\nu * - : \mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G) \to \mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$ sending λ to $\nu * \lambda$. Since I is a minimal left ideal, the image of $(\nu * -)|_I$ is contained in I. Since ν is definable, the map $(\nu * -)|_I : I \to I$ is continuous by Lemma 2.30. By Lemma 5.2, this map is also affine. By the Markov–Kakutani fixed-point theorem, there exists some $\mu \in I$ such that $\nu * \mu = \mu$. The "in particular" part of the statement follows since δ_g, $g \in G$ is a definable measure. □

5B. *Definably amenable groups.* We now shift our focus to the dividing line of definable amenability. We first describe all minimal left ideals in both $(\mathfrak{M}^{\mathrm{fs}}_{x}(\mathcal{G}, G), *)$ and $(\mathfrak{M}^{\mathrm{inv}}_{x}(\mathcal{G}, G), *)$ when $\mathcal{G}$ is definably amenable. We then make an observation about what happens outside of the definably amenable case. Recall that T is a complete NIP theory expanding a group, $\mathcal{G}$ is a monster model of T, G is a small elementary submodel of $\mathcal{G}$. The group $\mathcal{G}$ is *definably amenable* if there exists $\mu \in \mathfrak{M}_x(\mathcal{G})$ such that μ is $\mathcal{G}$-left-invariant.

Remark 5.14. (1) The group $\mathcal{G}$ is definably amenable if and only if for some $G' \models T$ there exists a G'-left-invariant $\mu \in \mathfrak{M}_x(G')$, if and only if for every $G' \models T$ there exists a G'-left-invariant $\mu \in \mathfrak{M}_x(G')$ (see [Hrushovski et al. 2008, Section 5]).

(2) If $G' \preceq \mathcal{G}$ and $\mu \in \mathfrak{M}_x(G')$ is G'-left-invariant, then the measure $\mu^{-1} \in \mathfrak{M}_x(G')$ defined by $\mu^{-1}(\varphi(x)) = \mu(\varphi(x^{-1}))$ for any $\varphi(x) \in \mathcal{L}_x(G')$ is G'-right-invariant, and vice versa. If $\mu \in \mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$, then also $\mu^{-1} \in \mathfrak{M}^{\dagger}_{x}(\mathcal{G}, G)$ (see [Chernikov and Simon 2018, Lemma 6.2]).

We will need the following fact.

Fact 5.15. *Assume that $\mathcal{G}$ is definably amenable and NIP.*

(i) [Chernikov et al. 2014, Proposition 3.5] *For any G-left-invariant measure $\mu_0 \in \mathfrak{M}_x(G)$* (which exists by Remark 5.14(1)) *there exists $\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ such that μ is $\mathcal{G}$-left-invariant and extends μ_0. The same holds for right-invariant measures by item* (2) *of Remark 5.14.*

(ii) [Chernikov et al. 2014, Theorem 3.17] *There exists $\nu \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ such that ν is G-left-invariant* (*but not necessarily $\mathcal{G}$-left-invariant*).

We remark that Fact 5.15(ii) follows from [Chernikov et al. 2014, Theorem 3.17] as $S_x^{\mathrm{fs}}(\mathcal{G}, G) = S_x(G^{\mathrm{ext}})$ (where G^{ext} is the Shelah's expansion of G by all externally definable subsets) and $\mathfrak{M}(S_x^{\mathrm{fs}}(\mathcal{G}, G)) = \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ (see Corollary 2.23). We now compute the minimal left ideals in definably amenable NIP groups, first in the finitely satisfiable case and then in the invariant case.

Proposition 5.16. *The group $\mathcal{G}$ is definably amenable if and only if $|I| = 1$ for some* (*equivalently*, *every*) *minimal left ideal I in $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$. And if $\mathcal{G}$ is definably amenable*, *then the minimal left ideals of $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ are precisely of the form $\{\nu\}$ for ν a G-left-invariant measure in $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$.*

Proof. Let I be a minimal left ideal, and assume that $I = \{\mu\}$. Then for any $g \in G$ we have $g \cdot \mu = \delta_g * \mu = \mu$, so μ is G-left-invariant. In particular, $\mu|_G$ is a G-left-invariant measure on $\mathfrak{M}_x(G)$, so $\mathcal{G}$ is definably amenable by Remark 5.14(1). And all minimal left ideals have the same cardinality by Fact 2.34(6).

Conversely, assume that $\mathcal{G}$ is definably amenable. By Fact 5.15(2) there exists some $\mu \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ such that μ is G-left-invariant. We claim that for any such μ, $\{\mu\}$ is a minimal left ideal of $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$. Let ν be any measure in $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$. Since ν is finitely satisfiable in G, by Lemma 2.25 there exists a net of measures in $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ of the form $(\mathrm{Av}(\bar{a}_i))_{i \in I}$ such that each $\bar{a}_i = (a_{i,1}, \dots, a_{i,n_i}) \in (G^x)^{n_i}$ for some $n_i \in \mathbb{N}$ and $\lim_{i \in I}(\mathrm{Av}(\bar{a}_i)) = \nu$. Fix any $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$. By the "moreover" part of Fact 2.29, the map $\lambda \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G) \mapsto (\lambda * \mu)(\varphi(x)) \in [0, 1]$ is continuous. Therefore,

$$\begin{aligned}(\nu * \mu)(\varphi(x)) &= \lim_{i \in I}\big((\mathrm{Av}(\bar{a}_i) * \mu)(\varphi(x))\big) \\ &= \lim_{i \in I}\left(\frac{1}{n_i}\sum_{j=1}^{n_i} \mu(\varphi(a_{i,j} \cdot x))\right) \overset{\text{(a)}}{=} \lim_{i \in I} \mu(\varphi(x)) = \mu(\varphi(x)).\end{aligned}$$

Equality (a) follows as μ is G-left-invariant and each $a_{i,j}$ is in G. It follows that $\nu * \mu = \mu$, and hence $\{\mu\}$ is a left ideal. $\square$

We now compute the minimal left ideals in the invariant case, but first we record an auxiliary lemma.

Lemma 5.17. *Assume that $f : S_x(G) \to [0, 1]$ is a Borel function. For any $b \in G$, we define the function $f \cdot b : S_x(G) \to [0, 1]$ via $(f \cdot b)(p) := f(p \cdot b)$* (recall Lemma 3.8). *If $\mu \in \mathfrak{M}_x(G)$ is G-right-invariant then*

$$\int_{S_x(G)} f\, d\mu = \int_{S_x(G)} (f \cdot b)\, d\mu.$$

Proof. For $b \in G$, consider the map $\gamma_b : S_x(G) \to S_x(G)$ defined by $\gamma_b(p) := p \cdot b$. The map γ_b is a continuous bijection. Hence we can consider the pushforward map $(\gamma_b)_* : \mathfrak{M}_x(G) \to \mathfrak{M}_x(G)$. Denote $(\gamma_b)_*(\mu)$ as μ_b. Fix a formula $\varphi(x) \in \mathcal{L}_x(G)$. We claim that $\gamma_b^{-1}([\varphi(x)]) = [\varphi(x \cdot b)]$.

We first show that $(\gamma_b)^{-1}([\varphi(x)]) = [\varphi(x \cdot b)]$. Assume that $p \in [\varphi(x \cdot b)]$. Then $\varphi(x) \in p \cdot b$ and so $p \cdot b \in [\varphi(x)]$. Hence $(\gamma_b)^{-1}(p \cdot b) \in (\gamma_b^{-1})([\varphi(x)])$. Since γ_b is a bijection, we have that $p = \gamma_b^{-1}(p \cdot b)$, which implies that $p \in (\gamma_b^{-1})([\varphi(x)])$. So $[\varphi(x \cdot b)] \subseteq (\gamma_b)^{-1}([\varphi(x)])$. Now assume that $p \in (\gamma_b)^{-1}([\varphi(x)])$. Then $\gamma_b(p) \in [\varphi(x)]$, and hence $p \cdot b \in [\varphi(x)]$, so $\varphi(x) \in p \cdot b$. By definition, $\varphi(x \cdot b) \in (p \cdot b) \cdot b^{-1}$, and since $(p \cdot b) \cdot b^{-1} = p$, we conclude that $\varphi(x \cdot b) \in p$. Hence $p \in [\varphi(x \cdot b)]$ and $(\gamma_b)^{-1}([\varphi(x)]) = [\varphi(x \cdot b)]$.

Now we show that $\mu_b = \mu$. Indeed, by G-right-invariance of μ and the previous paragraph we have

$$\mu_b(\varphi(x)) = \mu(\gamma_b^{-1}[\varphi(x)]) = \mu(\varphi(x \cdot b)) = \mu(\varphi(x)).$$

And so by Fact 2.1(iii) we have

$$\int_{S_x(G)} f\, d\mu = \int_{S_x(G)} f\, d\mu_b = \int_{S_x(G)} (f \circ \gamma_b)\, d\mu = \int_{S_x(G)} (f \cdot b)\, d\mu. \qquad \square$$

Proposition 5.18. *Assume that $\mathcal{G}$ is definably amenable. Let*

$$I_G^{\mathrm{inv}} := \{\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G) : \mu \text{ is } \mathcal{G}\text{-right-invariant}\}.$$

Then I_G^{inv} is a closed, nonempty, two-sided ideal. Moreover, I_G^{inv} is the unique minimal left ideal in $\mathfrak{M}_x(\mathcal{G}, G)$.

Proof. The set I_G^{inv} is closed since it is the complement of the union of basic open sets in $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$:

$$\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G) \setminus I_G^{\mathrm{inv}} = \bigcup_{\varphi(x) \in \mathcal{L}_x(\mathcal{G})} \bigcup_{s<t \in [0,1]} \bigcup_{g \in \mathcal{G}} \big(\{\mu : \mu(\varphi(x)) < s\} \cap \{\mu : \mu(\varphi(x \cdot g)) > t\}\big).$$

By Fact 5.15(1), we know that the set I_G^{inv} is nonempty. We first show that I_G^{inv} is a left ideal. Let $\mu \in I_G^{\mathrm{inv}}$ and $\nu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$. It suffices to show that the measure $\nu * \mu$ is $\mathcal{G}$-right-invariant. That is, we need to show that for any $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$ and $b \in \mathcal{G}$ we have $(\nu * \mu)(\varphi(x \cdot b)) = (\nu * \mu)(\varphi(x))$. Let $G' \prec \mathcal{G}$ be a small model containing G, b and the parameters of φ. For any $q \in S_y(G')$ and $a \models q$ in $\mathcal{G}$, letting $\varphi_b(x) := \varphi(x \cdot b)$ and noting that $a \cdot b \models q \cdot b$, we have

$$F^{\varphi_b'}_{\nu,G'}(q) = \nu(\varphi(x \cdot a \cdot b)) = F^{\varphi'}_{\nu,G'}(q \cdot b) = (F^{\varphi'}_{\nu,G'}(q)) \cdot b.$$

Hence, by Lemma 5.17,

$$\begin{aligned}(\nu * \mu)(\varphi(x \cdot b)) &= \int_{S_y(G')} F^{\varphi_b'}_{\nu}\, d\mu_{G'} \\ &= \int_{S_y(G')} ((F^{\varphi'}_{\nu}) \cdot b)\, d\mu_{G'} = \int_{S_y(G')} F^{\varphi'}_{\nu}\, d\mu_{G'} = (\nu * \mu)(\varphi(x)).\end{aligned}$$

We now argue that I^{inv}_G is a right ideal. Again let $\mu \in I^{\mathrm{inv}}_G$ and $\nu \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$, and fix $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$ and $G' \prec \mathcal{G}$ containing G and the parameters of φ. Using $\mathcal{G}$-right-invariance of μ, we have

$$(\mu * \nu)(\varphi(x)) = \int_{S_y(G')} F^{\varphi'}_{\mu}\, d\nu_{G'} = \int_{S_y(G')} \mu(\varphi(x))\, d\nu_{G'} = \mu(\varphi(x)).$$

Hence I^{inv}_G is a two-sided ideal.

Note that the previous computation shows that $\mu * \nu = \mu$ for any $\mu \in I^{\mathrm{inv}}_G$ and $\nu \in \mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$. So if J is any minimal left ideal of $\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$, then $I^{\mathrm{inv}}_G \subseteq J$. Since I^{inv}_G is two-sided, we have that $J \subseteq I^{\mathrm{inv}}_G$ (by Lemma 5.4). Hence $J = I^{\mathrm{inv}}_G$, and I^{inv}_G is the unique minimal left ideal. □

We recall some terminology and results from [Chernikov and Simon 2018] (switching from the action on the left to the action on the right everywhere).

Definition 5.19. (1) A type $p \in S_x(\mathcal{G})$ is *right f-generic* if for every $\varphi(x) \in p$ there is some small model $G \prec \mathcal{G}$ such that for any $g \in \mathcal{G}$, $\varphi(x \cdot g)$ does not fork over G.

(2) A type $p \in S_x(\mathcal{G})$ is *strongly right f-generic* if there exists some small $G \prec \mathcal{G}$ such that $p \cdot g \in S^{\mathrm{inv}}_x(\mathcal{G}, G)$ for all $g \in \mathcal{G}$. This is equivalent to the definition in [Chernikov and Simon 2018] since in NIP theories, a global type p does not fork over a model M if and only if p is M-invariant (see, e.g., [Hrushovski and Pillay 2011, Proposition 2.1]).

(3) Given a right f-generic p, let μ_p be defined via

$$\mu_p(\varphi(x)) := h\big(\{\pi(g) \in \mathcal{G}/\mathcal{G}^{00} : g \in \mathcal{G}, \varphi(x) \in p \cdot g\}\big),$$

where $\pi : \mathcal{G} \to \mathcal{G}/\mathcal{G}^{00}$ is the quotient map and $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$. Then $\mu_p \in \mathfrak{M}_x(\mathcal{G})$ and, assuming additionally that $\mathcal{G}$ is definably amenable, μ_p is $\mathcal{G}^{00}$-right-invariant (see [Chernikov and Simon 2018, Definition 3.16] for the details).

Fact 5.20. *Assume that $\mathcal{G}$ is definably amenable NIP.*

(1) *If $p \in S_x^{\mathrm{inv}}(\mathcal{G}, G)$ is right f-generic then p is strongly right f-generic over G and $\mu_p \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$. The set of all right f-generic types in $S_x(\mathcal{G})$ (and hence in $S_x^{\mathrm{inv}}(\mathcal{G}, G)$) is closed.*

(2) *Let $\mathfrak{I}(\mathcal{G})$ be the (closed convex) set of all $\mathcal{G}$-right-invariant measures in $\mathfrak{M}_x(\mathcal{G})$. Then the set $\mathrm{ex}(\mathfrak{I}(\mathcal{G}))$ of the extreme points of $\mathfrak{I}(\mathcal{G})$ is the set of all measures of the form μ_p for some right f-generic $p \in S_x(\mathcal{G})$.*

(3) *The map $p \mapsto \mu_p$ from the (closed) set of global right f-generic types to the (closed) set of global $\mathcal{G}$-right-invariant measures is continuous.*

Proof. (1) Any f-generic $p \in S_x^{\mathrm{inv}}(\mathcal{G}, G)$ is strongly f-generic over G by [Chernikov and Simon 2018, Proposition 3.9]. For any f-generic p, $\sup(\mu_p) \subseteq \overline{p \cdot \mathcal{G}}$, where $\overline{X}$ is the topological closure of X in $S_x(\mathcal{G})$ and $p \cdot \mathcal{G} = \{p \cdot g \in S_x(\mathcal{G}) : g \in \mathcal{G}\}$ is the orbit of p under the right action of $\mathcal{G}$ (by [Chernikov and Simon 2018, Remark 3.17(2)]). As p is strongly f-generic over G, we have $p \cdot \mathcal{G} \subseteq S_x^{\mathrm{inv}}(\mathcal{G}, G)$, and thus $\sup(\mu_p) \subseteq \overline{S_x^{\mathrm{inv}}(\mathcal{G}, G)} = S_x^{\mathrm{inv}}(\mathcal{G}, G)$. Hence $\mu_p \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ by Fact 2.22(2).

(2) This is [Chernikov and Simon 2018, Theorem 4.5].

(3) This is [Chernikov and Simon 2018, Proposition 4.3]. □

Adapting the proof of [Chernikov and Simon 2018, Theorem 4.5], we can describe the extreme points of the minimal ideal I_G^{inv}.

Corollary 5.21. *Assume that $\mathcal{G}$ is definably amenable NIP. Then*

(1) $\mathrm{ex}(I_G^{\mathrm{inv}}) = \{\mu_p : p \in S_x^{\mathrm{inv}}(\mathcal{G}, G)$ *is right f-generic*$\}$;

(2) $\mathrm{ex}(I_G^{\mathrm{inv}})$ *is a closed subset of I_G^{inv}, and I_G^{inv} is a Bauer simplex.*

Proof. If $p \in S_x^{\mathrm{inv}}(\mathcal{G}, G)$ is right f-generic, then μ_p is $\mathcal{G}$-right-invariant and $\mu_p \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ by Fact 5.20(1), so $\mu_p \in I_G^{\mathrm{inv}}$. By Fact 5.20(2), μ_p is extreme in $\mathfrak{I}(\mathcal{G})$, and thus, in particular, it is extreme in $I_G^{\mathrm{inv}} \subseteq \mathfrak{I}(\mathcal{G})$.

Conversely, assume that $\mu \in \mathrm{ex}(I_G^{\mathrm{inv}})$, and let

$$S := \{\mu_p : p \in S_x^{\mathrm{inv}}(\mathcal{G}, G) \text{ is right } f\text{-generic}\}.$$

Let $\overline{\mathrm{conv}}(S)$ be the closed convex hull of S. Then $\overline{\mathrm{conv}}(S) \subseteq I_G^{\mathrm{inv}}$ by Propositions 5.3 and 5.18. As μ is $\mathcal{G}$-right-invariant, by [Chernikov and Simon 2018, Lemma 3.26], for any $\varepsilon > 0$ and $\varphi_1(x), \dots, \varphi_k(x) \in \mathcal{L}_x(\mathcal{G})$, there exist some right f-generic $p_1, \dots, p_n \in \sup(\mu)$ such that $\mu(\varphi_j(x)) \approx_\varepsilon (1/n) \sum_{i=1}^n \mu_{p_i}(\varphi_j(x))$ for all $j \in [k]$. While [Chernikov and Simon 2018, Lemma 3.26] is stated for a single formula, it also applies to finitely many formulas by encoding them as appropriate instances of a single formula — formally, we apply [Chernikov and Simon 2018, Lemma 3.26] to the formula

$$\theta(x; y_0, \dots, y_k) := \bigvee_{i=1}^k (y_0 = y_i \wedge \varphi_k(x)).$$

As we have $p_i \in S^{\mathrm{inv}}_x(\mathcal{G}, G)$ for all $i \in [n]$, by Fact 2.22(2), it follows that $\mu \in \overline{\mathrm{conv}}(S)$, and it is still an extreme point of $\overline{\mathrm{conv}}(S) \subseteq I^{\mathrm{inv}}_G$. It follows that $\mu \in \bar{S}$, by the (partial) converse to the Krein–Milman theorem (see, e.g., [Chernikov and Simon 2018, Fact 4.1] applied to $C := \overline{\mathrm{conv}}(S)$). By Fact 5.20(3), the map $p \mapsto \mu_p$ from $S^{\mathrm{inv}}_x(\mathcal{G}, G)$ to $\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ is a continuous map from a compact to a Hausdorff space and thus also a closed map. It follows that $S = \bar{S}$, so $\mu \in S$.

By Corollary 2.23(2), we have an affine homeomorphism between $\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G)$ and $\mathcal{M}(S^{\mathrm{inv}}_x(\mathcal{G}, G))$, which restricts to an affine homeomorphism between I^{inv}_G and the set $\mathcal{M}_{\mathcal{G}}(S^{\mathrm{inv}}_x(\mathcal{G}, G))$ of all right-$\mathcal{G}$-invariant regular Borel probability measures on $S^{\mathrm{inv}}_x(\mathcal{G}, G)$. By Fact 2.37, $\mathcal{M}_{\mathcal{G}}(S^{\mathrm{inv}}_x(\mathcal{G}, G))$ is a Choquet simplex, so I^{inv}_G is a Bauer simplex (using Remark 2.40). □

Question 5.22. Can every Bauer simplex of the form $\mathcal{M}(X)$ with X a compact Hausdorff totally disconnected space be realized as a minimal left ideal of $(\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, G), *)$ for some definably amenable NIP group $\mathcal{G}$?

Example 5.23. Let $G := (\mathbb{R}; <, +)$, and let $\mathcal{G} \succ G$ be a monster model. As $\mathcal{G}$ is abelian, it is amenable as a discrete group and hence definably amenable. By Proposition 5.18, $\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G}, \mathbb{R})$ has a unique minimal left ideal I^{inv}_G. One checks directly that $p_{-\infty}$ (the unique type extending $\{x < a : a \in \mathcal{G}\}$) and $p_{+\infty}$ (the unique type extending $\{x > a : a \in \mathcal{G}\}$) are the right f-generics in $S^{\mathrm{inv}}_x(\mathcal{G}, G)$, and $\mu_{p_{+\infty}} = \delta_{p_{+\infty}}$, $\mu_{p_{-\infty}} = \delta_{p_{-\infty}}$. Hence, by Corollary 5.21, $|\mathrm{ex}(I^{\mathrm{inv}}_G)| = 2$ and

$$I^{\mathrm{inv}}_G = \{r\delta_{p_{+\infty}} + (1-r)\delta_{p_{-\infty}} : r \in [0, 1]\}.$$

(See also Example 6.21(1).)

Recall that $\mathcal{G}$ is *uniquely ergodic* if it admits a unique $\mathcal{G}$-left-invariant measure $\mu \in \mathfrak{M}_x(\mathcal{G})$ (see [Chernikov and Simon 2018, Section 3.4]). Recall that $\mathcal{G}$ is *fsg* if there exists a small $G \prec \mathcal{G}$ and $p \in S_x(\mathcal{G})$ such that $g \cdot p$ is finitely satisfiable in G for all $g \in \mathcal{G}$. All fsg groups are uniquely ergodic (see, e.g., [Simon 2015, Proposition 8.32]), but there exist uniquely ergodic NIP groups which are not fsg (see [Chernikov and Simon 2018, Remark 3.38]).

Corollary 5.24. (1) *If $\mathcal{G}$ is uniquely ergodic, then $I^{\mathrm{inv}}_G = \{\mu\}$, where μ is the unique $\mathcal{G}$-left-invariant measure.*

(2) *If $\mathcal{G}$ is, moreover, **fsg**, letting $\mu \in \mathfrak{M}_x(\mathcal{G})$ be the unique $\mathcal{G}$-left-invariant measure, $\{\mu\}$ is the unique minimal left ideal of $\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, G)$ (which is also two-sided).*

Proof. (1) For any $\mathcal{G}$-left-invariant measure μ, the measure μ^{-1} is $\mathcal{G}$-right-invariant (see Remark 5.14(2)), and vice versa. Moreover, from the definition, $\mu_1 = \mu_2$ if and only if $\mu_1^{-1} = \mu_2^{-1}$. It follows that if there exists a unique $\mathcal{G}$-left-invariant measure μ, then there exists a unique $\mathcal{G}$-right-invariant measure μ^{-1}. By [Chernikov and Simon

2018, Lemma 6.2] there also exists a measure ν which is simultaneously $\mathcal{G}$-left-invariant and $\mathcal{G}$-right-invariant. But then $\mu = \nu = \mu^{-1}$, so μ is also $\mathcal{G}$-right-invariant. And $\mu \in \mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ by Fact 5.15 and uniqueness, so $I_G^{\mathrm{inv}} = \{\mu\}$.

(2) By, e.g., [Simon 2015, Propositions 8.32, 8.33], $\mathcal{G}$ is fsg if and only if there exists a $\mathcal{G}$-left-invariant generically stable measure $\mu \in \mathfrak{M}_x(\mathcal{G})$, and then $\mathcal{G}$ is uniquely ergodic, so μ is also the unique $\mathcal{G}$-right-invariant measure. By Fact 5.15(i) and uniqueness of μ it follows that μ is invariant over G and hence generically stable over G (in fact, over an arbitrary small model). In particular, $\mu \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$, and it is the unique measure in $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ extending $\mu|_G$ (by [Hrushovski et al. 2013, Proposition 3.3]). Now assume that $\nu \in \mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$ is an arbitrary G-left-invariant measure. We have $\nu|_G = \mu|_G$, as by Fact 5.15(i) there exists some $\mathcal{G}$-left-invariant ν' extending $\nu|_G$, and thus $\nu' = \mu$, so $\nu|_G = \nu'|_G = \mu|_G$. But as μ is the unique measure in $\mathfrak{M}_x^{\mathrm{inv}}(\mathcal{G}, G)$ extending $\mu|_G$, it follows that $\nu = \mu$. If follows by Proposition 5.16 that $\{\mu\}$ is the unique minimal left ideal of $\mathfrak{M}_x^{\mathrm{fs}}(\mathcal{G}, G)$. Finally, in any semigroup, if the union of its minimal left ideals is nonempty, then it is a two-sided ideal [Clifford 1948]. Hence in our case $\{\mu\}$ is a two-sided ideal. □

Question 5.25. Can the fsg assumption be relaxed to unique ergodicity in Corollary 5.24(2)?

Our final observation in this section deals with nondefinably amenable groups.

Remark 5.26. Assume that $\mathcal{G}$ is not definably amenable. Let I be a minimal left ideal in $\mathfrak{M}^{\dagger}(\mathcal{G}, G)$. Then $\mathrm{ex}(I)$ is infinite.

Proof. For any $g \in G$, the map $\delta_g * - : \mathrm{ex}(I) \to \mathrm{ex}(I)$ is a bijection. Towards a contradiction, assume that $\mathrm{ex}(I)$ is finite, say $\mathrm{ex}(I) = \{\mu_1, \dots, \mu_n\}$. Consider the measure $\lambda \in \mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$ defined by $\lambda = \sum_{i=1}^{n} (1/n)\mu_i$. Then for any $g \in G$ we have $\delta_g * \lambda = \lambda$. Hence the measure $\lambda|_G$ is in $\mathfrak{M}_x(G)$ and is G-left-invariant. This contradicts the assumption that $\mathcal{G}$ is not definably amenable, by (1) and (2) of Remark 5.14. □

6. Constructing minimal left ideals

In this section, under some assumptions on the semigroup $(S_x^{\dagger}(\mathcal{G}, G), *)$ (applicable to some nondefinably amenable groups, e.g., $\mathrm{SL}_2(\mathbb{R})$), we construct a minimal left ideal of $(\mathfrak{M}_x^{\dagger}(\mathcal{G}, G), *)$ using a minimal left ideal and an ideal subgroup of $(S_x^{\dagger}(\mathcal{G}, G), *)$, and demonstrate that this minimal left ideal is parametrized by a space of regular Borel probability measures over a compact Hausdorff space.

6A. ***Basic lemmas.*** We will need some auxiliary lemmas connecting convolution and left ideals. We assume that $T = \mathrm{Th}(\mathcal{G})$ is NIP throughout.

Lemma 6.1. *Let $\mu, \nu \in \mathfrak{M}_x^{\dagger}(\mathcal{G}, G)$. If $\mu * \delta_p = \mu$ for every $p \in \sup(\nu)$, then $\mu * \nu = \mu$.*

Proof. Fix a formula $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$. Let $G' \prec \mathcal{G}$ be a small model containing G and the parameters of φ. We have

$$(\mu * \nu)(\varphi(x)) = \int_{\sup(\nu|_{G'})} F^{\varphi'}_{\mu,G'}\, d(\nu_{G'}).$$

By Fact 2.22, $\sup(\nu)$ is a subset of $S^{\dagger}_x(\mathcal{G}, G)$. For any $q \in \sup(\nu)$ we have $F^{\varphi'}_{\mu,G'}(q) = \mu(\varphi(x \cdot b)) = (\mu * \delta_p)(\varphi(x)) = \mu(\varphi(x))$, where $b \models q$. Hence

$$\int_{\sup(\nu|_{G'})} F^{\varphi'}_{\mu,G'}\, d(\nu_{G'}) = \int_{S_y(G')} \mu(\varphi(x))\, d(\nu_{G'}) = \mu(\varphi(x)),$$

so $\mu * \nu = \mu$. □

Lemma 6.2 (T is NIP). *Assume that I is a left ideal of $(S^{\dagger}_x(\mathcal{G}, G), *)$. Then $\mathfrak{M}(I)$* (see Definition 2.16) *is a left ideal of* $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$.

Proof. Let $p \in S^{\dagger}_x(\mathcal{G}, G)$ and $\mu \in \mathfrak{M}(I)$. We first argue that $\delta_p * \mu \in \mathfrak{M}(I)$. Assume towards a contradiction that $\delta_p * \mu \notin \mathfrak{M}(I)$. Then there exists some $q \in \sup(\delta_p * \mu)$ such that $q \notin I$. Then there exists $\psi(x) \in \mathcal{L}_x(\mathcal{G})$ such that $\psi(x) \in q$ and $[\psi(x)] \cap I = \varnothing$. Since $\psi(x) \in q$ and $q \in \sup(\delta_p * \mu)$, we have $(\delta_p * \mu)(\psi(x)) > 0$. Let now $G' \prec \mathcal{G}$ be a small model containing G and the parameters of ψ. Then

$$(\delta_p * \mu)(\psi(x)) = \int F^{\psi'}_{\delta_p,G'}\, d(\mu_{G'}) > 0,$$

so there exists some $t \in \sup(\mu|_{G'})$ such that $F^{\psi'}_{\delta_p,G'}(t) = 1$. Fix $\hat{t} \in \sup(\mu)$ such that $\hat{t}|_{G'} = t$ (which exists by, e.g., [Chernikov and Gannon 2022, Proposition 2.8]), and since $\mu \in \mathfrak{M}(I)$ we have $\hat{t} \in \operatorname{supp}(\mu) \subseteq I \subseteq S^{\dagger}_x(\mathcal{G}, G)$. Unpacking the notation, we conclude that $\psi(x) \in p * \hat{t}$. Since $\hat{t} \in I$, it also follows that $p * \hat{t} \in I$. Hence $[\psi(x)] \cap I \neq \varnothing$, a contradiction.

Now letting $\nu \in \mathfrak{M}^{\dagger}_x(\mathcal{G}, G)$, we want to show that $\nu * \mu \in \mathfrak{M}(I)$. By Lemma 2.25, we have that $\nu = \lim_{i \in I} \operatorname{Av}(\bar{p}_i)$ for some net I, with $\bar{p}_i = (p_{i,1}, \ldots, p_{i,n_i}) \in I^{n_i}$, $n_i \in \mathbb{N}$ for each $i \in I$. By left continuity of convolution (Fact 2.29) we have

$$\nu * \mu = \lim_{i \in I} (\operatorname{Av}(\bar{p}_i) * \mu) = \lim_{i \in I} \left(\frac{1}{n_i} \sum_{j=1}^{n_i} (\delta_{p_j} * \mu) \right).$$

By the previous paragraph $\delta_{p_{j_i}} * \mu \in \mathfrak{M}(I)$ for each $i \in I$. Then by convexity of $\mathfrak{M}(I)$ (Lemma 2.17), also $\operatorname{Av}(\bar{p}_i) * \mu \in \mathfrak{M}(I)$ for each $i \in I$. Since $\mathfrak{M}(I)$ is closed (again, Lemma 2.17), $\nu * \mu = \lim_{i \in I}(\operatorname{Av}(\bar{p}_i) * \mu) \in \mathfrak{M}(I)$. Therefore $\mathfrak{M}(I)$ is a left ideal. □

Remark 6.3. We remark that minimality of the left ideal need not be preserved in Lemma 6.2. Indeed, let $G := (S^1, \cdot, ^{-1}, C(x, y, z))$ be the standard unit circle group over $\mathbb{R}$, with C the cyclic clockwise ordering, and let T_O be the corresponding theory. If $\mathcal{G}$ is a monster model of T_O, then the semigroup $(S^{\mathrm{fs}}_x(\mathcal{G}, S^1), *)$ has a unique proper (and hence minimal) left ideal $I := S^{\mathrm{fs}}_x(\mathcal{G}, S^1) \setminus \{\mathrm{tp}(a/\mathcal{G}) : a \in S^1\}$. Let λ be the Keisler measure corresponding to the normalized Haar measure on S^1. The measure λ is smooth and right-invariant; in particular, $\mathcal{G}$ is fsg (see [Chernikov and Gannon 2022, Example 4.2] and [Simon 2015, Proposition 8.33]). By Lemma 6.2, $\mathfrak{M}(I)$ is a left ideal of $(\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, G), *)$. Note that $\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, S^1)$ contains a unique minimal left ideal $\{\lambda\}$ by Corollary 5.24(2), and $\{\lambda\} \subsetneq \mathfrak{M}(I)$ since the latter contains δ_p for every global type p finitely satisfiable in S^1 but not realized in it.

We now recall how the ideal subgroups act on a minimal left ideal. The following is true in any compact left topological semigroup; we include a proof for completeness in our setting.

Corollary 6.4. *Let I be a minimal left ideal in $S^{\dagger}_x(\mathcal{G}, G)$ and u an idempotent in I. Let p be any element in I. Then the map $(- * p)|_{u*I} : u * I \to u * I$ is a continuous bijection. Moreover, $(- * p)|_{u*I} = (- * (u * p))|_{u*I}$.*

Proof. We have $(u * I) * p = u * (I * p) = u * I$ as $I * p = I$ by Fact 2.34(5) (using Fact 2.28).

To show surjectivity, fix $r \in u * I$; as $u * p \in u * I$ and $u * I$ is a group with identity u, there exists some $s \in u * I$ such that $s * (u * p) = u$; then $r * s \in u * I$, and $(r * s) * p = (r * s * u) * p = r * (s * (u * p)) = r * u = r$. To show injectivity, assume $r * p = t * p$ for some $r, t \in u * I$; as also $r * u = r$ and $t * u = t$, we have $r * (u * p) = t * (u * p)$, and therefore, taking inverses in the group $u * I$, we have $r * (u * p) * (u * p)^{-1} = t * (u * p) * (u * p)^{-1}$, so $r * u = t * u$, so $r = t$. Finally, the map is continuous as a restriction of a continuous map $- * p : S^{\dagger}_x(\mathcal{G}, G) \to S^{\dagger}_x(\mathcal{G}, G)$. The "moreover" part follows directly from associativity. □

6B. *Compact ideal subgroups* (CIG1). We define CIG1 semigroups and show that under this assumption, we can describe a minimal left ideal of the semigroup of measures.

Definition 6.5. We say that the semigroup $(S^{\dagger}_x(\mathcal{G}, G), *)$ is CIG1 (or "admits compact ideal subgroups") if there exists some minimal left ideal I and idempotent $u \in I$ such that $u * I$ is a compact group with the induced topology from I. We let h_{u*I} denote the normalized Haar measure on $u * I$, and define the Keisler measure $\mu_{u*I} \in \mathfrak{M}_x(\mathcal{G})$ as follows:

$$\mu_{u*I}(\varphi(x)) := h_{u*I}([\varphi(x)] \cap u * I).$$

Remark 6.6. Suppose that $(S^{\dagger}_x(\mathcal{G}, G), *)$ is CIG1. Then any minimal left ideal witnesses this property, i.e., for any minimal left ideal J of $S^{\dagger}_x(\mathcal{G}, G)$ there exists an idempotent $v \in J$ such that $v * J$ is a compact group with the induced topology.

Proof. Suppose $(S^{\dagger}_x(\mathcal{G}, G), *)$ is CIG1. Fix a minimal left ideal I and an idempotent u in I such that $u * I$ is a compact group. Let J be any other minimal left ideal.

By Fact 2.34(6) there exists an idempotent $v \in J$ such that $u * v = v$, $v * u = u$, and the map $(- * v)|_I : I \to J$ is a homeomorphism mapping $u * I$ to $v * J$. Note that the restriction to $u * I$ is a group homomorphism (indeed, for $p_1, p_2 \in u * I$, $(p_1 * v) * (p_2 * v) = p_1 * v * u * p_2 * v = p_1 * u * p_2 * v = (p_1 * p_2) * v$) and hence a continuous group isomorphism. Since it is also a homeomorphism onto its range $v * J$, as the restriction of a homeomorphism, it follows that $v * J$ is a compact group. $\square$

Lemma 6.7. *The semigroup $(S^{\dagger}_x(\mathcal{G}, G), *)$ is* CIG1 *if either of the following holds*:

(1) *For some minimal left ideal I, every $p \in I$ is definable.*

(2) *The ideal group of $S^{\dagger}_x(\mathcal{G}, G)$ is finite.*

Proof. (1) Fix $p \in I$ and let $u \in I$ be the unique idempotent such that $p \in u * I$ (by Fact 2.34(4)). Since p is definable, the map $(p * -)|_I : I \to I$ is continuous (by Lemma 2.30) and hence also closed. Since I is compact, the image of $(p * -)|_I$ is compact and is equal to $u * I$. Hence $(u * I, *)$ is a compact Hausdorff space, an abstract group, and both left multiplication and right multiplication are continuous. By Fact 2.35, $(u * I, *)$ is a compact group.

(2) This is obvious. $\square$

Example 6.8. (1) Let $G := (\mathbb{Z}, +, <)$, and consider the sets

$$I^+ := \{q \in S^{\mathrm{inv}}_x(\mathcal{G}, \mathbb{Z}) : (a < x) \in q \text{ for all } a \in \mathcal{G}\},$$
$$I^- := \{q \in S^{\mathrm{inv}}_x(\mathcal{G}, \mathbb{Z}) : (x < a) \in q \text{ for all } a \in \mathcal{G}\}.$$

Then $I := I^+ \cup I^-$ is the unique minimal left ideal of $(S^{\mathrm{inv}}_x(\mathcal{G}, \mathbb{Z}), *)$. Note that every type in I is definable (over $\mathbb{Z}$). By Lemma 6.7, the semigroup $(S^{\mathrm{inv}}_x(\mathcal{G}, \mathbb{Z}), *)$ is CIG1. The ideal subgroups are $(I^-, *)$ and $(I^+, *)$, both isomorphic to $\widehat{\mathbb{Z}}$ as topological groups.

(2) Consider $G := \mathrm{SL}_2(\mathbb{R})$ as a definable subgroup in $(\mathbb{R}, \cdot, +)$. If I is a minimal left ideal of $\big(S^{\mathrm{fs}}_x(\mathcal{G}, \mathrm{SL}_2(\mathbb{R})), *\big)$ and u is an idempotent in I, then $u * I \cong \mathbb{Z}/2\mathbb{Z}$ by [Gismatullin et al. 2015, Theorem 3.17], so the semigroup is CIG1. Note that $\mathrm{SL}_2(\mathbb{R})$ is not definably amenable [Hrushovski et al. 2008, Remark 5.2; Conversano and Pillay 2012, Lemma 4.4(1)].

(3) There exist fsg groups that are not CIG1. Consider the circle group from Remark 6.3. The minimal left ideal of $(S^{\mathrm{fs}}(\mathcal{G}, S^1), *)$ is precisely $S^{\mathrm{fs}}(\mathcal{G}, S^1)$. As in (1), this left ideal can be decomposed into two ideal subgroups as follows. Let $\mathrm{st} : \mathcal{G} \to S^1$ be the standard part map. Consider the sets

$$I^R := \{q \in S^{\mathrm{fs}}_x(\mathcal{G}, G) : \text{if } b \models q, \text{ then } C(\mathrm{st}(b), b, a) \text{ for any } a \in S^1\},$$
$$I^L := \{q \in S^{\mathrm{fs}}_x(\mathcal{G}, G) : \text{if } b \models q, \text{ then } C(a, b, \mathrm{st}(b)) \text{ for any } a \in S^1\}.$$

Then both I^R and I^L are ideal subgroups which are isomorphic (as abstract groups) to S^1, and $S^{\mathrm{fs}}_x(\mathcal{G}, S^1) = I^R \sqcup I^L$. Moreover, I^R and I^L are dense subsets of $S_x^{\mathrm{fs}}(\mathcal{G}, S^1)$. Note that if I^R were compact (with the induced topology), we would have $I^R = S_x^{\mathrm{fs}}(\mathcal{G}, S^1)$, a contradiction. The same argument applies to I^L. Therefore, $\big(S_x^{\mathrm{fs}}(\mathcal{G}, \mathrm{SL}_2(\mathbb{R})), *\big)$ is not CIG1.

Lemma 6.9. *Assume that* $(S_x^{\dagger}(\mathcal{G}, G), *)$ *is* CIG1. *Let* $I \subseteq S_x^{\dagger}(\mathcal{G}, G)$ *be a minimal left ideal and* u *an idempotent in* I *such that* $u * I$ *is a compact group. Then for any* $p \in u * I$ *we have* $\mu_{u*I} * \delta_p = \mu_{u*I}$ *and* $\delta_p * \mu_{u*I} = \mu_{u*I}$.

Proof. Fix $p \in u * I$ and $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$. Let $G' \prec \mathcal{G}$ be a small model containing G and the parameters of φ. Let $a \models p|_{G'}$, and let p^{-1} be the unique element of the group $u * I$ such that $p * p^{-1} = u$.

Claim 1: $$(\mu_{u*I} * \delta_p)(\varphi(x)) = \mu_{u*I}(\varphi(x)).$$

Proof. We have the following computation, using right-invariance of the Haar measure h_{u*I} on $u * I$:

$$
\begin{aligned}
&(\mu_{u*I} * \delta_p)(\varphi(x)) \\
&\qquad = \int_{S_y(G')} F^{\varphi'}_{\mu_{u*I}}\, d(\delta_p|_{G'}) = F^{\varphi'}_{\mu_{u*I}}(p|_{G'}) = \mu_{u*I}(\varphi(x \cdot a)) \\
&\qquad = h_{u*I}\big([\varphi(x \cdot a)] \cap u * I\big) = h_{u*I}\big(\{q \in u * I : \varphi(x \cdot a) \in q\}\big) \\
&\qquad = h_{u*I}\big(\{q \in u * I : \varphi(x) \in q * p\}\big) = h_{u*I}\big(\{q \in u * I : \varphi(x) \in q\} * p^{-1}\big) \\
&\qquad = h_{u*I}\big(\{q \in u * I : \varphi(x) \in q\}\big) = \mu_{u*I}(\varphi(x)).
\end{aligned}
$$
□

Claim 2: $$(\delta_p * \mu_{u*I})(\varphi(x)) = \mu_{u*I}(\varphi(x)).$$

Proof. Let $r : S_y(\mathcal{G}) \to S_y(G')$ be the restriction map. Let $\tilde{\mu}_{u*I}$ be the extension of μ_{u*I} to a regular Borel probability measure on $S_x(\mathcal{G})$. By construction, $\mathrm{supp}(\tilde{\mu}_{u*I}) = \sup(\mu_{u*I}) = u * I$ and $\tilde{\mu}_{u*I}|_{u*I} = h_{u*I}$. Using left-invariance of h_{u*I} we have

$$
\begin{aligned}
&(\delta_p * \mu_{u*I})(\varphi(x)) \\
&\quad = \int_{S_y(G')} F^{\varphi'}_{\delta_p}\, d(\mu_{u*I}|_{G'}) = \int_{S_y(\mathcal{G})} (F^{\varphi'}_{\delta_p} \circ r)\, d\mu_{u*I} \\
&\quad = \tilde{\mu}_{u*I}\big(\{q \in S_x(\mathcal{G}) : (F^{\varphi'}_{\delta_p} \circ r)(q) = 1\}\big) = \tilde{\mu}_{u*I}\big(\{q \in S_x(\mathcal{G}) : \varphi(x) \in p * q\}\big) \\
&\quad = \tilde{\mu}_{u*I}\big(\{q \in u * I : \varphi(x) \in p * q\}\big) = h_{u*I}\big(p^{-1} * \{q \in u * I : \varphi(x) \in q\}\big) \\
&\quad = h_{u*I}\big(\{q \in u * I : \varphi(x) \in q\}\big) = \mu_{u*I}(\varphi(x)).
\end{aligned}
$$
□

Hence the statement holds. □

Lemma 6.10. *Assume that* $(S_x^{\dagger}(\mathcal{G}, G), *)$ *is* CIG1. *Let* $I \subseteq S_x^{\dagger}(\mathcal{G}, G)$ *be a minimal left ideal and* u *an idempotent in* I *such that* $u * I$ *is a compact group. Then for any* $p \in I$ *we have* $\mu_{u*I} * \delta_p = \mu_{u*I}$.

Proof. For any $p \in I$ we have

$$\mu_{u*I} * \delta_p = (\mu_{u*I} * \delta_u) * \delta_p = \mu_{u*I} * (\delta_u * \delta_p) = \mu_{u*I} * \delta_{u*p} = \mu_{u*I},$$

where the first and the last equalities are by Lemma 6.9, as $u, u * p \in u * I$. □

Theorem 6.11. *Assume* $(S^{\dagger}_x(\mathcal{G}, G), *)$ *is* CIG1. *Let* $I \subseteq S^{\dagger}_x(\mathcal{G}, G)$ *be a minimal left ideal and* u *an idempotent in* I *such that* $u * I$ *is a compact group. Then* $\mathfrak{M}(I) * \mu_{u*I}$ *is a minimal left ideal of* $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$, *containing an idempotent* μ_{u*I}.

Proof. We first argue that μ_{u*I} is an element of some minimal left ideal of $\mathfrak{M}^{\dagger}_x(\mathcal{G}, G)$. We know that $\mathfrak{M}(I)$ is a closed (by Fact 2.34 and Lemma 2.17) left ideal of $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$ (by Lemma 6.2). Hence there exists some $L \subseteq \mathfrak{M}(I)$ such that L is a minimal left ideal of $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$, and we show that $\mu_{u*I} \in L$. Let $\nu \in \mathfrak{M}(I)$ be arbitrary. If $p \in \sup(\nu)$, then $p \in I$. By Lemma 6.10, we then have $\mu_{u*I} * \delta_p = \mu_{u*I}$ for every $p \in \sup(\nu)$. By Lemma 6.1 this implies $\mu_{u*I} * \nu = \mu_{u*I}$, and therefore $\mu_{u*I} * \mathfrak{M}(I) = \{\mu_{u*I}\}$. In particular, $\mu_{u*I} * L = \{\mu_{u*I}\}$, and since L is a left ideal this implies $\mu_{u*I} \in L$ (and also that μ_{u*I} is an idempotent).

Then $\mathfrak{M}^{\dagger}_x(\mathcal{G}, G) * \mu_{u*I} = L$ by Fact 2.34(5). We also have that $L * \mu_{u*I} = L$ since $\mu_{u*I} \in L$ and L is a minimal left-ideal. Thus

$$L = L * \mu_{u*I} \subseteq \mathfrak{M}(I) * \mu_{u*I} \subseteq \mathfrak{M}^{\dagger}_x(\mathcal{G}, G) * \mu_{u*I} = L.$$

Hence $\mathfrak{M}(I) * \mu_{u*I} = L$, so $\mathfrak{M}(I) * \mu_{u*I}$ is a minimal left ideal of $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$. □

Corollary 6.12. *Suppose that* $(S^{\dagger}_x(\mathcal{G}, G), *)$ *is* CIG1. *Let* I *be a minimal left ideal and* u *an idempotent in* I *such that* $u * I$ *is a compact group. Let* J *be any minimal left ideal of* $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$. *Then* J *and* $\mathfrak{M}(I) * \mu_{u*I}$ *are affinely homeomorphic.*

Proof. By Fact 2.34(6), Lemma 5.2, and Theorem 6.11. □

6C. ***Compact ideal subgroups in minimal ideals with Hausdorff quotients* (CIG2).** In this section we define CIG2 semigroups and show that under this stronger assumption, any minimal left ideal of $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$ is affinely homeomorphic to the space of regular Borel probability measures over a certain compact Hausdorff space given by a quotient of a minimal left ideal in $(S^{\dagger}_x(\mathcal{G}, G), *)$.

Definition 6.13. Let I be a minimal left ideal in $(S^{\dagger}_x(\mathcal{G}, G), *)$. We define the quotient space $K_I := I/\sim$, where $p \sim q$ if and only if p and q are elements of the same ideal subgroup of I, i.e., there exists some idempotent $u \in I$ such that $p, q \in u * I$. We endow K_I with the induced quotient topology and write elements of K as $[u * I]$, where u is an idempotent in I.

The quotient topology on K_I is automatically compact, but may not be Hausdorff. CIG2 stipulates that this quotient is Hausdorff.

Definition 6.14. We say that the semigroup $(S^{\dagger}_x(\mathcal{G}, G), *)$ is CIG2 if there exists a minimal left ideal I such that

(i) for any idempotent $u \in I$, $u * I$ is compact;

(ii) for any $p \in I$ and $u' \in \mathrm{id}(I)$, the map $(p * -)|_{u' * I}$ is continuous (note that the range of this map is $u * I$, where $u \in \mathrm{id}(I)$ is such that $p \in u * I$);

(iii) K_I is Hausdorff.

We remark that in the above definition, (i) follows from (iii) since each $u * I$ is a preimage of a point (and hence a closed set) in K_I under the quotient map.

Lemma 6.15.[2] *The semigroup* $(S^{\dagger}_x(\mathcal{G}, G), *)$ *is* CIG2 *if either of the following holds*:

(1) *The ideal group of* $(S^{\dagger}_x(\mathcal{G}, G), *)$ *is finite.*

(2) *For some minimal ideal* $I \subseteq S^{\dagger}_x(\mathcal{G}, G)$*, every* $p \in I$ *is definable.*

Proof. (1) Assume that the ideal group of $(S^{\dagger}_x(\mathcal{G}, G), *)$ is finite. Then the first two conditions of CIG2 are clearly satisfied, and we show (iii) from Definition 6.14. Suppose that I is a minimal left ideal in $(S^{\dagger}_x(\mathcal{G}, G), *)$, and let u be an idempotent in I. Let us denote elements of $u * I$ as $\mathfrak{g}$. Then $u * I$ acts on I on the right via $p \cdot \mathfrak{g} := p * \mathfrak{g}$, and the orbit equivalence relation under this group action is the same as the equivalence relation $\sim$ in the definition of K_I. Indeed, u is the identity of $u * I$ and $p * u = p$ for all $p \in I$ by Fact 2.34(2); if $p \cdot \mathfrak{g} = q$ and $p \in u' * I$ for some $u' \in \mathrm{id}(I)$, then $q = p * \mathfrak{g} \in (u' * I) * \mathfrak{g} = u' * (I * \mathfrak{g}) \subseteq u' * I$; and conversely, if $p, q \in u' * I$, using that $u' * I$ is a group and Fact 2.34(2), we have $p = (q * q^{-1}) * p = q * (q^{-1} * p) = q * (u' * r) = q * (u' * u) * r = (q * u') * (u * r) = q * (u * r) = q \cdot \mathfrak{g}$ for some $r \in I$ and $\mathfrak{g} := u * r$. This action is continuous by left continuity of convolution.

So $K_I = I/(u * I)$, and the quotient of any Hausdorff space by a continuous finite group action remains Hausdorff. Hence K_I is Hausdorff.

(2) The conditions (i) and (ii) of CIG2 hold since every type in the minimal left ideal I is definable, as in the proof of Lemma 6.7(1). Let $u \in I$ be an idempotent. Arguing as in (1) we get $K_I = I/(u * I)$. The right action of the group $u * I$ on I is continuous on the right, and by the assumption and Lemma 2.30 it is also continuous on the left and therefore continuous by the Ellis joint continuity theorem (Fact 2.35). Thus $I/(u * I)$ is Hausdorff, as the quotient of a Hausdorff space by the continuous action of a compact group. $\square$

The next fact follows directly from the definitions and Fact 2.35.

Remark 6.16. If $(S^{\dagger}_x(\mathcal{G}, G), *)$ is CIG2, then it is CIG1. Moreover, if I is a minimal left ideal of $(S^{\dagger}_x(\mathcal{G}, G), *)$ witnessing CIG2, then for any idempotent $u \in I$, $u * I$ is

[2]We thank the referee for pointing out a more general version of Lemma 6.15, as well as Remark 6.17.

a compact group with the induced topology. Thus for every idempotent u in I, the measure μ_{u*I} is well defined.

Remark 6.17. (1) In the proof of Lemma 6.15(2), it suffices to assume that for some idempotent $u \in I$, $u * I$ is closed and that for all $p \in I$, the map $p * -|_{u*I}$ is continuous.

(2) We also have the following equivalence: CIG2 holds if and only if CIG1 holds, and the map $u' * -|_{u*I}$ is continuous for some u witnessing CIG1 and every idempotent $u' \in I$.

Indeed, since $u * I$ is compact, it follows that each $u' * I$ is compact, and thus closed and $u' * -|_{u*I}$ is a homeomorphism. Since it is also a group isomorphism, each $u' * I$ is a compact group. Now, given any $p \in u' * I$, we have $p = u' * p = u' * u * p$, so left multiplication by p of elements of $u * I$ is the composition of left multiplication by $u * p \in u * I$ (continuous since $u * I$ is a topological group) and left multiplication by u' (continuous by assumption), and therefore it is continuous and we conclude by (1).

Example 6.18. Both examples (1) and (2) from Example 6.8 are CIG2.

(1) The semigroup $(S_x^{\mathrm{inv}}(\mathcal{G}, \mathbb{Z}), *)$ is CIG2 by Lemma 6.15(2) as all types in I are definable (note that we have $|K_I| = 2$).

(2) The ideal group of $\big(S_x^{\mathrm{fs}}(\mathcal{G}, \mathrm{SL}_2(\mathbb{R})), *\big)$ is finite ($\cong \mathbb{Z}/2\mathbb{Z}$), so it is CIG2 by Lemma 6.15(1).

Lemma 6.19. *Assume that $(S_x^{\dagger}(\mathcal{G}, G), *)$ is* CIG2, *and let I be a minimal left ideal witnessing it. Then for any $p \in I$ and $u \in \mathrm{id}(I)$ we have $\delta_p * \mu_{u*I} = \mu_{u'*I}$, where u' is the unique idempotent in I such that $p \in u' * I$.*

Proof. Fix $u, u' \in \mathrm{id}(I)$. Then the transition map $\rho_{u,u'} := (u' * -)|_{u*I} : u * I \to u' * I$ is an isomorphism of topological groups (it is a group isomorphism by Fact 2.34(3) and continuous by (ii) in CIG2, and $\rho_{u',u} \circ \rho_{u,u'} = \mathrm{id}_{u*I}$). Let $\Phi_{u,u'} : \mathcal{M}(u * I) \to \mathcal{M}(u' * I)$ be the corresponding pushforward map. Note that $\Phi_{u',u} \circ \Phi_{u,u'} = \mathrm{id}_{\mathcal{M}(u*I)}$. Moreover, $\Phi_{u,u'}(h_{u*I}) = h_{u'*I}$ because $\Phi_{u,u'}(h_{u*I})$ is a regular Borel probability measure on $u' * I$ which is right-invariant, and this property characterizes the normalized Haar measure. By a computation similar to the proof of Claim 2 in Lemma 6.9, for any $\varphi(x) \in \mathcal{L}_x(\mathcal{G})$ we have

$$\begin{aligned}
(\delta_u * \mu_{u'*I})(\varphi(x)) &= h_{u'*I}\big(\{q \in u' * I : \varphi(x) \in u * q\}\big) \\
&= (\Phi_{u,u'}(h_{u*I}))\big(\{q \in u' * I : \varphi(x) \in u * q\}\big) \\
&= h_{u*I}\big(\rho_{u,u'}^{-1}\big(\{q \in u' * I : \varphi(x) \in u * q\}\big)\big) \\
&= h_{u*I}\big(u * \{q \in u' * I : \varphi(x) \in u * q\}\big) \\
&= h_{u*I}\big(\{q \in u * I : \varphi(x) \in q\}\big) = \mu_{u*I}(\varphi(x)),
\end{aligned}$$

and hence $\delta_u * \mu_{u'*I} = \mu_{u*I}$. Now let $p \in u' * I$. By Lemma 6.9 and the above computation, using that $p = p * u'$ by Fact 2.34(2), we have

$$\delta_p * \mu_{u*I} = \delta_{p*u'} * \mu_{u*I} = (\delta_p * \delta_{u'}) * \mu_{u*I} = \delta_p * (\delta_{u'} * \mu_{u*I}) = \delta_p * \mu_{u'*I} = \mu_{u'*I}. \quad \square$$

Theorem 6.20. *Suppose that* $(S^{\dagger}_x(\mathcal{G}, G), *)$ *is* CIG2. *Let* $I \subseteq S^{\dagger}_x(\mathcal{G}, G)$ *be a minimal left ideal witnessing* CIG2. *Then all minimal left ideals of* $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$ *are affinely homeomorphic to* $\mathcal{M}(K_I)$ (*in particular*, *they are Bauer simplices by item* (2) *of Fact 2.41*).

Proof. Let $u \in \mathrm{id}(I)$. By Remark 6.16 and Corollary 6.12, it suffices to show that $\mathfrak{M}(I) * \mu_{u*I} \cong \mathcal{M}(K_I)$. For ease of notation, denote the minimal left ideal $\mathfrak{M}(I) * \mu_{u*I}$ as L. Let $q : I \to K_I$ denote the (continuous) quotient map, and $q_* : \mathcal{M}(I) \to \mathcal{M}(K_I)$ the corresponding pushforward map. Note that q_* is affine by Fact 2.1(iii). By Proposition 2.21, we have an affine homeomorphism $\gamma : \mathfrak{M}(I) \to \mathcal{M}(I)$. Let $\Phi := (q_* \circ \gamma)|_L$. We claim that Φ is an affine homeomorphism. Note that Φ is the restriction of the composition of two continuous affine maps, so Φ itself is a continuous affine map. It suffices to show that Φ is a bijection (since it is automatically a closed map as L is compact and $\mathcal{M}(K_I)$ is Hausdorff by Fact 2.1(i) as K_I is compact Hausdorff by CIG2).

Claim 1: Φ is surjective.

Proof. The extreme points of $\mathcal{M}(K_I)$ are the Dirac measures concentrating on the elements of K_I (see, e.g., [Simon 2011, Example 8.16]). By the Krein–Milman theorem, the set

$$\left\{ \sum_{i=1}^{n} r_i \delta_{[u_i * I]} : [u_i * I] \in K_I, r_i \in \mathbb{R}_{>0}, \sum_{i=1}^{n} r_i = 1, n \in \mathbb{N} \right\}$$

is dense in $\mathcal{M}(K_I)$. Fix some $u_1, \dots, u_n \in \mathrm{id}(I)$ and $r_1, \dots, r_n \in \mathbb{R}_{>0}$ such that $\sum_{i=1}^n r_i = 1$. It suffices to find some $\mu \in L$ such that $\Phi(\mu) = \sum_{i=1}^n r_i \delta_{[u_i*I]}$ (as Φ is a closed map, it will follow that $\Phi(L) = \mathcal{M}(K_I)$).

Let $\lambda := \sum_{i=1}^n r_i \delta_{u_i} \in \mathfrak{M}^{\dagger}_x(\mathcal{G}, G)$. Since $\mu_{u*I} \in L$ (by Theorem 6.11) and L is a left ideal, also $\lambda * \mu_{u*I} \in L$. By Lemmas 5.2 and 6.19, we have

$$\lambda * \mu_{u*I} = \left(\sum_{i=1}^{n} r_i \delta_{u_i} \right) * \mu_{u*I} = \sum_{i=1}^{n} r_i (\delta_{u_i} * \mu_{u*I}) = \sum_{i=1}^{n} r_i \mu_{u_i*I},$$

and as γ and q_* are affine this implies

$$\Phi(\lambda * \mu_{u*I}) = \Phi\left(\sum_{i=1}^{n} r_i \mu_{u_i*I} \right) = \sum_{i=1}^{n} r_i q_*(\tilde{\mu}_{u_i*I}) = \sum_{i=1}^{n} r_i \delta_{[u_i*I]},$$

where $\tilde{\mu}_{u_i*I} \in \mathcal{M}(I)$ is the unique regular Borel probability measure extending μ_{u_i*I}, i.e., $\tilde{\mu}_{u_i*I}(X) = h_{u_i*I}(X \cap u_i * I)$ for any Borel $X \subseteq I$, where h_{u_i*I} is the Haar measure on $u_i * I$. Hence Φ is surjective. $\square$

Claim 2: Φ is injective.

Proof. Suppose that λ and ν are in L and $\lambda \neq \nu$. It suffices to find a continuous function $f : K_I \to \mathbb{R}$ such that

$$\int_{K_I} f\, d(\Phi(\lambda)) \neq \int_{K_I} f\, d(\Phi(\nu)).$$

Since $\lambda \neq \nu$, there exists some $\psi(x) \in \mathcal{L}_x(\mathcal{G})$ such that $\lambda(\psi(x)) \neq \nu(\psi(x))$. Consider the function $f_\psi : I \to \mathbb{R}$ defined via $f_\psi(p) := (\delta_p * \mu_{u*I})(\psi(x))$. This map is continuous since the map $(- * \mu_{u*I})(\psi(x)) : \mathfrak{M}^\dagger_x(\mathcal{G}, G) \to \mathbb{R}$ is continuous by the "moreover" part of Fact 2.29 (and the map $p \in S^\dagger_x(\mathcal{G}, G) \mapsto \delta_p \in \mathfrak{M}^\dagger_x(\mathcal{G}, G)$ is continuous). Moreover, f_ψ factors through q. Indeed, assume that $q(p_1) = q(p_2)$ for some $p_1, p_2 \in I$. Then there exists some $w \in \mathrm{id}(I)$ such that $p_1, p_2 \in w * I$. Then by Lemma 6.19 we have

$$f_\psi(p_1) = (\delta_{p_1} * \mu_{u*I})(\psi(x)) = \mu_{w*I}(\psi(x)) = (\delta_{p_2} * \mu_{u*I})(\psi(x)) = f_\psi(p_2).$$

By the universal property of quotient maps, there exists a unique continuous function $f : K_I \to \mathbb{R}$ such that $f_\psi = f \circ q$. Since $\lambda \in L \subseteq \mathfrak{M}(I)$ (by the proof of Theorem 6.11), by Lemma 2.25 there exists a net of measures $(\mathrm{Av}(\bar{p}_j))_{j\in J}$ such that $\bar{p}_j = (p_{j,1}, \dots, p_{j,n_j}) \in I^{n_j}$ and $\lim_{j\in J} \mathrm{Av}(\bar{p}_j) = \lambda$ for each $j \in J$. Because γ is an affine homeomorphism, we then have $\gamma(\lambda) = \lim_{j\in J}\big((1/n_j)\sum_{k=1}^{n_j} \delta_{p_{j,k}}\big)$. Hence we have the following computation:

$$\begin{aligned}
&\int_{K_I} f\, d(\Phi(\lambda)) \\
&= \int_{K_I} f\, d\big(q_*(\gamma(\lambda))\big) = \int_I (f\circ q)\, d(\gamma(\lambda)) = \int_I f_\psi\, d(\gamma(\lambda)) \\
&= \int_{S_x(\mathcal{G})} f_\psi\, d(\gamma(\lambda)) = \int_{S_x(\mathcal{G})} f_\psi\, d\bigg(\lim_{j\in J}\bigg(\frac{1}{n_j}\sum_{k=1}^{n_j} \delta_{p_{j,k}}\bigg)\bigg) \\
&= \lim_{j\in J} \int_{S_x(\mathcal{G})} f_\psi\, d\bigg(\frac{1}{n_j}\sum_{k=1}^{n_j} \delta_{p_{j,k}}\bigg) = \lim_{j\in J}\big(\mathrm{Av}(\bar{p}_j) * \mu_{u*I}(\psi(x))\big) \qquad \text{(by Fact 2.1(ii))} \\
&= \big((\lim_{j\in J} \mathrm{Av}(\bar{p}_j)) * \mu_{u*I}\big)(\psi(x)) = (\lambda * \mu_{u*I})(\psi(x)) = \lambda(\psi(x)),
\end{aligned}$$

where the last equality holds by Fact 2.34(2), as μ_{u*I} is an idempotent in L. A similar computation shows that $\int_{K_I} f\, d(\Phi(\nu)) = \nu(\psi(x)) \neq \lambda(\psi(x))$, so Φ is injective. □

Claims 1 and 2 establish the theorem. □

Example 6.21. (1) Let $G := (\mathbb{R}, +, <)$. Then the semigroup $S^{\mathrm{inv}}_x(\mathcal{G}, \mathbb{R})$ is CIG2. Indeed, the unique minimal left ideal of $S^{\mathrm{inv}}_x(\mathcal{G}, \mathbb{R})$ is $I = \{p_{-\infty}, p_{+\infty}\}$, and both elements of I are idempotents (see Example 4.5(3)). The ideal subgroups of I

are $\{p_{-\infty}\}$ and $\{p_{+\infty}\}$, both of which are obviously compact groups under induced topology. We have $\mathfrak{M}(I)=\{r\delta_{p_{-\infty}}+(1-r)\delta_{p_{+\infty}} : r\in[0,1]\}$, and if $u=p_{\pm\infty}$ then $\mu_{u*I}=\delta_{p_{\pm\infty}}$.

Now we let $\nu\in\mathfrak{M}(I)$, and then $\nu=r\delta_{p_{-\infty}}+s\delta_{p_{+\infty}}$ for some $r,s\in[0,1]$ with $r+s=1$. Then $\nu*\mu_{p_{\pm\infty}*I}=(r\delta_{p_{-\infty}}+s\delta_{p_{+\infty}})*\mu_{p_{\pm\infty}*I}=(r\delta_{p_{-\infty}}+s\delta_{p_{+\infty}})*\delta_{p_{\pm\infty}}=r(\delta_{p_{-\infty}}*\delta_{p_{\pm\infty}})+s(\delta_{p_{+\infty}}*\delta_{p_{\pm\infty}})=r\delta_{p_{-\infty}}+s\delta_{p_{+\infty}}$. Therefore $\mathfrak{M}(I)*\mu_{p_{\pm\infty}*I}=\mathfrak{M}(I)$, and so $\mathfrak{M}(I)*\mu_{p_{\pm\infty}*I}\cong\mathcal{M}(\{0,1\})$ is a minimal ideal of $(\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G},\mathbb{R}),*)$.

(2) Let $G:=(\mathbb{Z},+,<)$. Then the semigroup $S^{\mathrm{inv}}_x(\mathcal{G},\mathbb{Z})$ is CIG2, the unique minimal left ideal of $S^{\mathrm{inv}}_x(\mathcal{G},\mathbb{Z})$ is $I=I^+\sqcup I^-$ and the ideal subgroups of I are I^+ and I^- (see Examples 6.8 and 6.18). Both ideal subgroups are compact groups under induced topology, isomorphic to $\hat{\mathbb{Z}}$ as a topological group.

Let $u^+\in I^+$ and $u^-\in I^-$ be the identity group elements in I^- and I^+, respectively. Then $\mu_{u^\pm*I}$ is the Haar measure on $I^\pm\cong\hat{\mathbb{Z}}$. For every $\nu\in\mathfrak{M}(I)$ we can write $\nu=r\nu^-+s\nu^+$ for the measures ν^- and ν^+ defined by

$$\nu^-(\varphi(x))=\frac{\nu(\varphi(x)\wedge x<b)}{\nu(x<b)},\quad \nu^+(\varphi(x))=\frac{\nu(\varphi(x)\wedge x>c)}{\nu(x>c)}\quad\text{and}\quad b<\mathbb{Z}<c.$$

We also have $\nu*\mu_{u^\pm*I}=(r\nu^-+s\nu^+)*\mu_{u^\pm*I}=r(\nu^-*\mu_{u^\pm*I})+s(\nu^+*\mu_{u^\pm*I})=r\mu_{u^-*I}+s\mu_{u^+*I}$. Therefore $\mathfrak{M}(I)*\mu_{u_{\pm\infty}*I}=\{r\mu_{u^-*I}+s\mu_{u^+*I} : r+s=1\}\cong\mathcal{M}(\{0,1\})$ is a minimal ideal of $(\mathfrak{M}^{\mathrm{inv}}_x(\mathcal{G},\mathbb{R}),*)$.

Fact 6.22 [Gismatullin et al. 2015]. *Let $\mathbb{R}\prec\mathcal{R}$ be a saturated real closed field, $G:=\mathrm{SL}_2(\mathbb{R})$ and $\mathcal{G}:=\mathrm{SL}_2(\mathcal{R})$. Consider the definable subgroups of $\mathcal{G}$ given by*

$$\mathcal{T}:=\left\{\left(\begin{bmatrix}x&-y\\y&x\end{bmatrix}\right):x^2+y^2=1\right\}\quad\text{and}\quad\mathcal{H}:=\left\{\left(\begin{bmatrix}b&c\\0&b^{-1}\end{bmatrix}\right):b\in\mathcal{R}_{>0},c\in\mathcal{R}\right\}.$$

Let $p_0:=\mathrm{tp}((b,c)/\mathbb{R})$ such that $b>\mathbb{R}$ and $c>\mathrm{dcl}(\mathbb{R}\cup\{b\})$. We view p_0 as a type in $S_{\mathcal{H}}(\mathbb{R})$[3] *identifying (b,c) with the matrix $\left[\begin{smallmatrix}b&c\\0&b^{-1}\end{smallmatrix}\right]$. Let $q_0:=\mathrm{tp}((x,y)/\mathbb{R})$, where y is positive infinitesimal and $x>0$ is the positive square root of $1-y^2$. We view q_0 as a type in $S_{\mathcal{T}}(\mathbb{R})$ identifying (x,y) with the matrix $\left[\begin{smallmatrix}x&-y\\y&x\end{smallmatrix}\right]$. We let r_0 be $\mathrm{tp}(t\cdot h/\mathbb{R})\in S_{\mathcal{G}}(\mathbb{R})$, where $h\in\mathcal{H}$ realizes p_0 and $t\in\mathcal{T}$ realizes the unique coheir of q_0 over $\mathbb{R}\cup\{h\}$. Then*

(1) *$S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R},\mathbb{R})*r_0$ is a minimal left ideal of $S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R},\mathbb{R})$;*

(2) *any ideal subgroup of $S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R},\mathbb{R})*r_0$ is isomorphic to $\mathbb{Z}/2\mathbb{Z}$; in particular, if we let r_1 be the unique element in $S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R},\mathbb{R})*r_0$ such that $r_1*r_1=r_0$ and $r_1\neq r_0$, then $\{r_0,r_1\}$ is an ideal subgroup.*

[3] As usual, we denote by $S_{\mathcal{H}}(-)$ the space of types concentrating on the definable set $\mathcal{H}$; all of our results can be modified in an obvious manner to apply to definable groups in an arbitrary theory, as opposed to theories expanding a group.

Example 6.23. Let $\mathcal{G} = \mathrm{SL}_2(\mathcal{R})$ and $S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R}, \mathbb{R})$ be the collection of global types concentrated on $\mathcal{G}$ which are finitely satisfiable in $\mathrm{SL}_2(\mathbb{R})$. By Fact 6.22, $\{r_0, r_1\}$ is an ideal subgroup of $S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R}, \mathbb{R})$ which is trivially a compact group with the induced topology, and $\frac{1}{2}(\delta_{r_0} + \delta_{r_1})$ is the normalized Haar measure on it. By Theorem 6.11, $\mathfrak{M}(S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R}, \mathbb{R}) * r_0) * \frac{1}{2}(\delta_{r_0} + \delta_{r_1})$ is a minimal left ideal in $\mathfrak{M}^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R}, \mathbb{R})$. Moreover, this minimal left ideal is affinely homeomorphic to $\mathcal{M}(K_{S^{\mathrm{fs}}_{\mathcal{G}}(\mathcal{R},\mathbb{R}) * r_0})$ by Theorem 6.20 (see the notation there), which is a Bauer simplex with infinitely many extreme points (by Remark 5.26).

More generally, we have:[4]

Remark 6.24. If $\mathcal{G}$ is NIP, not definably amenable and $(S^{\dagger}_x(\mathcal{G}, G), *)$ is CIG2, then the quotient K_I is infinite for each minimal ideal I in $(S^{\dagger}_x(\mathcal{G}, G), *)$, and the minimal ideals in $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$ are Bauer simplices, each with infinitely many extreme points (by Fact 2.41(2), Remark 5.26 and Theorem 6.20).

Remark 6.25. Assume that $\mathcal{G}$ is NIP and $(S^{\mathrm{fs}}_x(\mathcal{G}, G), *)$ is CIG2. Then the following are equivalent:

(1) $\mathcal{G}$ is definably amenable.

(2) $|K_I| = 1$ for each minimal left ideal I in $S^{\mathrm{fs}}_x(\mathcal{G}, G)$.

(3) K_I is finite for some minimal left ideal I in $S^{\mathrm{fs}}_x(\mathcal{G}, G)$.

Proof. (1) $\Rightarrow$ (2) By definable amenability and Proposition 5.16, $|J| = 1$ for every minimal left ideal J in $(\mathfrak{M}^{\mathrm{fs}}_x(\mathcal{G}, G), *)$. By Theorem 6.20, J is affinely homeomorphic to $\mathcal{M}(K_I)$ for some minimal left ideal I of $(S^{\mathrm{fs}}_x(\mathcal{G}, G), *)$ and therefore $|K_I| = 1$ also. By Fact 2.34(6), we have $|K_{I'}| = 1$ for every minimal left ideal I' of $(S^{\mathrm{fs}}_x(\mathcal{G}, G), *)$.

(2) $\Rightarrow$ (3) This is trivial.

(3) $\Rightarrow$ (1) This is by Remark 6.24 applied for $\dagger = \mathrm{fs}$. □

Remark 6.26. The implication (1) $\Rightarrow$ (2) in Remark 6.25 does not hold when $(S^{\mathrm{fs}}_x(\mathcal{G}, G), *)$ is replaced by $(S^{\mathrm{inv}}_x(\mathcal{G}, G), *)$. Indeed, $(\mathbb{Z}, +, <)$ is NIP, definably amenable, CIG2, but $|K_I| = 2$ (see Example 6.18(1)).

Question 6.27. It would be interesting to describe minimal left ideals of the semigroup $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$ for some nondefinably amenable groups $\mathcal{G}$ where a description of the minimal left ideals/ideal subgroups of $(S^{\dagger}_x(\mathcal{G}, G), *)$ is known (other than $\mathrm{SL}_2(\mathbb{R})$), including certain algebraic groups definable in $\mathbb{Q}_p$ [Penazzi et al. 2019; Bao and Yao 2022] or in certain dp-minimal fields [Jagiella 2021].

Question 6.28. Is the set of extreme points of a minimal left ideal of $(\mathfrak{M}^{\dagger}_x(\mathcal{G}, G), *)$ always closed, or at least Borel, in a (not necessarily definably amenable) NIP group $\mathcal{G}$?

[4]We thank the referee for suggesting the following two remarks.

Acknowledgements. We are very grateful to the referee for a particularly detailed and insightful report and many valuable suggestions on improving the paper.

We thank Sergei Starchenko and Anand Pillay for helpful conversations. Both authors were partially supported by the NSF CAREER grant DMS-1651321, and Chernikov was additionally supported by a Simons fellowship.

References

[Abodayeh and Murphy 1997] K. Abodayeh and G. J. Murphy, "Compact topological semigroups", *Proc. Roy. Irish Acad. Sect. A* **97**:2 (1997), 131–137. MR Zbl

[Alfsen 1971] E. M. Alfsen, *Compact convex sets and boundary integrals*, Ergeb. Math. Grenzgeb. **57**, Springer, 1971. MR Zbl

[Bao and Yao 2022] J. Bao and N. Yao, "Definably topological dynamics of p-adic algebraic groups", *Ann. Pure Appl. Logic* **173**:4 (2022), art. id. 103077, 17 pages. MR Zbl

[Bhaskara Rao and Bhaskara Rao 1983] K. P. S. Bhaskara Rao and M. Bhaskara Rao, *Theory of charges: a study of finitely additive measures*, Pure and Applied Mathematics **109**, Academic Press, New York, 1983. MR Zbl

[Chernikov and Gannon 2022] A. Chernikov and K. Gannon, "Definable convolution and idempotent Keisler measures", *Israel J. Math.* **248**:1 (2022), 271–314. MR Zbl

[Chernikov and Simon 2018] A. Chernikov and P. Simon, "Definably amenable NIP groups", *J. Amer. Math. Soc.* **31**:3 (2018), 609–641. MR Zbl

[Chernikov et al. 2014] A. Chernikov, A. Pillay, and P. Simon, "External definability and groups in NIP theories", *J. Lond. Math. Soc.* (2) **90**:1 (2014), 213–240. MR Zbl

[Clifford 1948] A. H. Clifford, "Semigroups containing minimal ideals", *Amer. J. Math.* **70** (1948), 521–526. MR Zbl

[Cohen and Collins 1959] H. Cohen and H. S. Collins, "Affine semigroups", *Trans. Amer. Math. Soc.* **93** (1959), 97–113. MR Zbl

[Conant et al. 2021] G. Conant, K. Gannon, and J. Hanson, "Keisler measures in the wild", preprint, 2021. arXiv 2103.09137

[Conversano and Pillay 2012] A. Conversano and A. Pillay, "Connected components of definable groups and o-minimality, I", *Adv. Math.* **231**:2 (2012), 605–623. MR Zbl

[Dales et al. 2016] H. G. Dales, F. K. Dashiell, Jr., A. T.-M. Lau, and D. Strauss, *Banach spaces of continuous functions as dual spaces*, Springer, 2016. MR Zbl

[Ellis 1957] R. Ellis, "Locally compact transformation groups", *Duke Math. J.* **24** (1957), 119–125. MR Zbl

[Ellis et al. 2001] D. B. Ellis, R. Ellis, and M. Nerurkar, "The topological dynamics of semigroup actions", *Trans. Amer. Math. Soc.* **353**:4 (2001), 1279–1320. MR Zbl

[Gannon 2019] K. Gannon, "Local Keisler measures and NIP formulas", *J. Symb. Log.* **84**:3 (2019), 1279–1292. MR Zbl

[Gismatullin et al. 2015] J. Gismatullin, D. Penazzi, and A. Pillay, "Some model theory of $\mathrm{SL}(2, \mathbb{R})$", *Fund. Math.* **229**:2 (2015), 117–128. MR Zbl

[Glasner 2007] E. Glasner, "Enveloping semigroups in topological dynamics", *Topology Appl.* **154**:11 (2007), 2344–2363. MR Zbl

[Holický and Spurný 2003] P. Holický and J. Spurný, "Perfect images of absolute Souslin and absolute Borel Tychonoff spaces", *Topology Appl.* **131**:3 (2003), 281–294. MR Zbl

[Hrushovski and Pillay 2011] E. Hrushovski and A. Pillay, "On NIP and invariant measures", *J. Eur. Math. Soc.* **13**:4 (2011), 1005–1061. MR Zbl

[Hrushovski et al. 2008] E. Hrushovski, Y. Peterzil, and A. Pillay, "Groups, measures, and the NIP", *J. Amer. Math. Soc.* **21**:2 (2008), 563–596. MR Zbl

[Hrushovski et al. 2013] 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

[Jagiella 2021] G. Jagiella, "Topological dynamics and NIP fields", *Ann. Pure Appl. Logic* **172**:9 (2021), art. id. 103010, 15 pages. MR Zbl

[Kawada and Itô 1940] Y. Kawada and K. Itô, "On the probability distribution on a compact group, I", *Proc. Phys.-Math. Soc. Japan* (3) **22** (1940), 977–998. MR Zbl

[Lawson 1974] J. D. Lawson, "Joint continuity in semitopological semigroups", *Illinois J. Math.* **18** (1974), 275–285. MR Zbl

[Penazzi et al. 2019] D. Penazzi, A. Pillay, and N. Yao, "Some model theory and topological dynamics of p-adic algebraic groups", *Fund. Math.* **247**:2 (2019), 191–216. MR Zbl

[Phelps 2001] R. R. Phelps, *Lectures on Choquet's theorem*, 2nd ed., Lecture Notes in Mathematics **1757**, Springer, 2001. MR Zbl

[Pillay 2004] A. Pillay, "Type-definability, compact Lie groups, and o-minimality", *J. Math. Log.* **4**:2 (2004), 147–162. MR Zbl

[Shelah 2008] S. Shelah, "Minimal bounded index subgroup for dependent theories", *Proc. Amer. Math. Soc.* **136**:3 (2008), 1087–1091. MR Zbl

[Simon 2011] B. Simon, *Convexity: an analytic viewpoint*, Cambridge Tracts in Mathematics **187**, Cambridge Univ. Press, 2011. MR Zbl

[Simon 2015] P. Simon, *A guide to NIP theories*, Lecture Notes in Logic **44**, Association for Symbolic Logic, Chicago, 2015. MR Zbl

[Stromberg 1959] K. Stromberg, "A note on the convolution of regular measures", *Math. Scand.* **7** (1959), 347–352. MR Zbl

[Wendel 1954] J. G. Wendel, "Haar measure and the semigroup of measures on a compact group", *Proc. Amer. Math. Soc.* **5** (1954), 923–929. MR Zbl

Received 26 Feb 2022. Revised 26 Jul 2022.

ARTEM CHERNIKOV:

chernikov@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, United States

KYLE GANNON:

gannon@math.ucla.edu
Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, United States

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.233

Higher amalgamation properties in measured structures

David M. Evans

Using an infinitary version of the hypergraph removal lemma due to Towsner, we prove a model-theoretic higher amalgamation result. In particular, we obtain an independent amalgamation property which holds in structures that are measurable in the sense of Macpherson and Steinhorn, but which is not generally true in structures that are supersimple of finite SU-rank. We use this to show that some of Hrushovski's non-locally-modular, supersimple ω-categorical structures are not MS-measurable.

1. Introduction

Towsner [2018] gives an infinitary version of the hypergraph removal lemma (quoted as Theorem 2.3 here), stated as a rather general measure-theoretic result. We use this to prove a model-theoretic higher amalgamation result (Theorem 2.4), again in the presence of a definable measure. In particular, we obtain an independent amalgamation property (Corollary 3.2; quoted below as Corollary 1.1) which holds in structures that are measurable in the sense of Macpherson and Steinhorn.

The statement of this independent amalgamation property makes no mention of measure and it makes sense in any supersimple structure of finite SU-rank. However, it is not generally true in structures which are supersimple of finite SU-rank. In Theorem 4.7, we use a Hrushovski construction to produce a structure which is ω-categorical, supersimple of SU-rank 1 and which does not satisfy the conclusion of Corollary 3.2. It follows that this structure is not MS-measurable.

The question of whether any (nontrivial) ω-categorical Hrushovski construction can be MS-measurable is open, and this is an important special case of the more general question of whether ω-categorical MS-measurable structures are necessarily one-based. Paolo Marimon [2022a; 2022b] has used a different and more generally applicable approach to show that a much wider class of ω-categorical, supersimple Hrushovski constructions are not MS-measurable. It is also unknown whether any of the ω-categorical Hrushovski constructions can be pseudofinite. In Remarks 4.6 we note that, as a by-product of our approach to non-MS-measurability, we obtain

MSC2020: primary 03C45; secondary 03C13.
Keywords: model theory, amalgamation properties, measurability, pseudofiniteness, Hrushovski constructions.

information about what coarse pseudofinite dimension would have to be in such a structure, if it were pseudofinite.

We begin with a rough outline of what we mean by a "higher amalgamation property". This is adapted to the form of the Towsner's paper, so is slightly different from other presentations (for example in [Hrushovski 2012]).

Suppose L is a first-order language and $\mathcal{M}$ is an L-structure with domain M and $C \subseteq M$. Let T denote the theory of $\mathcal{M}$. We will assume that $\mathcal{M}$ is "large" (for example $\aleph_1$-saturated, if L is countable) and C has smaller cardinality than that of M. Suppose $n \geq 2$ is a natural number. In an *n-amalgamation problem* over C we are looking for an n-tuple $\bar{b} = (b_1, \ldots, b_n)$ which satisfies certain constraints on subtuples $\bar{b}_I = (b_i : i \in I)$ with $I \subseteq [n] = \{1, \ldots, n\}$ of size $n-1$. The constraints should be in terms of the parameters C, say in the form of satisfying a type, or partial type, $\Phi_I(\bar{x}_I)$, over C. Here, $\bar{x} = (x_1, \ldots, x_n)$ is an n-tuple of variables and $x_I = (x_i : i \in I)$. So, subject to reasonable compatibility requirements, such as $\Phi_I(\bar{x}_I)$ and $\Phi_J(\bar{x}_J)$ having the same restriction to $\bar{x}_{I\cap J}$, we are looking for a solution $\bar{b} \models \bigwedge_I \Phi_I(\bar{x}_I)$, or, in terms of the sets $A_I = \{\bar{a} \in M^n : \mathcal{M} \models \Phi_I(\bar{a})\}$, an element of $\bigcap_I A_I$. If the Φ_I are complete types over C, we might refer to this as a *type-amalgamation* problem.

There are well-known variations on this. If $\mathcal{M}$ carries a notion of independence (or dimension on definable sets) then in an *independent* n-amalgamation problem over C, we are also looking for the b_i to be independent over C. Of course, in this case, the individual constraints $\Phi_I(\bar{x}_I)$ should have solutions which are independent over C. For example, if T is stable, then for all n, any independent type-amalgamation problem over a model (with n complete types over the model) has a solution. If T is simple, then this is true for $n = 2, 3$ (the case $n = 3$ is of course the independence theorem of Kim and Pillay). However, there are examples of supersimple theories of finite SU-rank which do not have independent 4-amalgamation over a model.

Our main result, Theorem 2.4, is an n-amalgamation property which holds in a general context where the set of n-tuples from which we are looking for a solution carries a well-behaved probability measure (see Section 2A for a precise statement). The general form of the statement is that we assume there our n-amalgamation problem has "degenerate" solutions $\bar{b} = (b_1, \ldots, b_n)$, where the b_i are interalgebraic over C. The conclusion is that the set of all solutions is of positive measure (and in particular, there are solutions where the b_i are not interalgebraic). Of course, for this to work, we need to ensure that there are enough solutions to the Φ_I: in the above notation, we require that the measure of A_I is positive, for each $(n-1)$-set I.

If $\mathcal{M}$ is an MS-measurable structure (see Section 3B for definitions and background) there is a strong interaction between dimension and measure. The structure $\mathcal{M}$ is supersimple of finite SU-rank and each definable subset has an associated dimension (which can be taken as SU-rank for the purposes of this introduction). Each definable set also carries a (definable) probability measure on its definable subsets with the property that a subset has positive measure if and only if it has the same dimension as the ambient definable set.

From Theorem 2.4 we obtain the following independent amalgamation result (Corollary 3.2), which holds in any MS-measurable M.

Corollary 1.1. *Suppose $\mathcal{M}$ is an MS-measurable structure and $S_1, \dots, S_n$ are infinite C-definable sets, for some finite $C \subset M$. Let $S = S_1 \times \cdots \times S_n$ and for $I \subset [n] = \{1, \dots, n\}$, let $\pi_I : S \to \prod_{i \in I} S_i$ be the projection map. Suppose $E \subseteq S$ is a C-definable subset such that*

(a) *if $I \subset [n]$ and $|I| = n - 1$, then $\dim(\pi_I(E)) = \sum_{i \in I} \dim(S_i)$, and*

(b) *if $(b_1, \dots, b_n) \in E$, then $b_i \in \operatorname{acl}(C \cup \{b_j : j \neq i\})$.*

Then

$$\dim\big\{\bar{b} \in S : \pi_I(\bar{b}) \in \pi_I(E) \text{ for all } I \text{ with } |I| = n - 1\big\} = \dim(S).$$

Note that this does not tell us anything if $\mathcal{M}$ has trivial algebraic closure. Note also that it does not refer to the measure, so it makes sense in any supersimple theory (more properly, any S_1-theory) of finite SU-rank. In Section 4 we give an example of a supersimple structure of SU-rank 1 which does not satisfy the above result: so we have an independent amalgamation result which holds in MS-measurable structures, but which is not generally true in finite rank supersimple structures.

This paper is a revised version of some unpublished notes written in 2011–2012. The original version made use of Towsner's unpublished article [2010] and proved Theorem 2.4 under stronger assumptions on the definability of the measure and the behaviour of the measure under projection maps with finite fibres. In 2019, I sent a copy of the notes to Ehud Hrushovski, who observed that these assumptions could be weakened. He also gave examples of additional contexts in which the weaker assumptions would hold: see Section 3C here.

Towsner's published paper [2018] contains a reworking of [Towsner 2010] which involves a weaker assumption on the definability of the measure. In revising the original notes, I have therefore rewritten the proof of Theorem 2.4 to follow the approach and notation of [Towsner 2018].

The structure of the paper is as follows. In Section 2A we give the necessary notation and background to state Towsner's version of the hypergraph removal lemma from [Towsner 2018]. In Section 2B we deduce the main result, Theorem 2.4, from this. Our result is related to a standard deduction of Szemerédi's theorem from the hypergraph removal lemma: we make this explicit in Section 3A. In Section 3B, we discuss MS-measurability and prove Corollary 3.2, stated above. Additional examples in NIP theories are mentioned briefly in Section 3C. In Section 4, we discuss the ω-categorical Hrushovski constructions and their relationship to various open questions around MS-measurable ω-categorical structures. The main result of the section is Theorem 4.7, where we construct an ω-categorical structure which is of SU-rank 1 and which does not satisfy the amalgamation property in Corollary 3.2.

2. An amalgamation theorem for measured structures

2A. *Measured structures.* The following setup is taken from Towsner's paper [2018]. Chapter 1 of [Kallenberg 1997] is a convenient reference for the basic measure theory we need.

We work with a structure $\mathcal{M}$ with domain M. The following notation is introduced in [Towsner 2018, Section 2]. If V is a finite set of indices, then a V-tuple from M is a function $\bar{a}_V : V \to M$ and we denote the set of these by M^V. A V-tuple of variables will generally be denoted by $\bar{x}_V$. If $I \subseteq V$ then $\bar{a}_I \in M^I$ is the restriction of this to I. If U and W are disjoint sets, we write $\bar{a}_U \cup \bar{a}_W$ for the $(U \cup W)$-tuple extending $\bar{a}_U$ and $\bar{a}_W$. If $B \subseteq M^{U \cup W}$ and $\bar{a}_W \in M^W$, then $B(\bar{a}_W)$ denotes the fibre (or "slice") $\{\bar{a}_U \in M^U : \bar{a}_U \cup \bar{a}_W \in B\}$.

In what follows, V is a fixed finite set of indices $V = \{1, \dots, n\} = [n]$ for some $n \in \mathbb{N}$. We often denote $\bar{a}_V$ or $\bar{x}_V$ simply by $\bar{a}$ or $\bar{x}$, dropping the reference to V.

Definition 2.1 [Towsner 2018, Definition 4.1]. Suppose that for each $U \subseteq V$ we have a Boolean algebra $\mathcal{B}^0_U$ of subsets of M^U such that

- $\varnothing \in \mathcal{B}^0_U$;
- $\mathcal{B}^0_U \times \mathcal{B}^0_W \subseteq \mathcal{B}^0_{U \cup W}$ for disjoint $U, W \subseteq V$;
- if $U, W \subseteq V$ are disjoint, $\bar{a}_W \in M^W$ and $B \in \mathcal{B}^0_{U \cup W}$, then $B(\bar{a}_W) \in \mathcal{B}^0_U$.

For $I \subseteq V$ we define $\mathcal{B}^0_{V,I}$ to be the Boolean algebra generated by subsets

$$\{\bar{a}_V \in M^V : \bar{a}_I \in B\},$$

where $B \in \mathcal{B}^0_I$.

In all cases we will drop the superscript 0 to indicate the σ-algebra generated by the Boolean algebra.

The main result we need from [Towsner 2018] is Theorem 2.3 below. When we use this, $\mathcal{B}^0_V$ will consist of the parameter-definable subsets of M^V, so the reader may assume this from now on. We then refer to the elements of $\mathcal{B}_V$ as Borel sets. If $X \subseteq M$, then $\mathcal{B}^0_V(X)$ will denote the X-definable subsets of M^V, and we use a corresponding variation in the notation for the algebras introduced above. We will assume sufficient saturation, so that it makes sense to identify a formula defining a Borel set with its solution set in M. In particular, if the language is countable, we assume that $\mathcal{M}$ is $\aleph_1$-saturated. If the model is multisorted, then we can restrict each variable to having values in a particular sort.

Suppose, with the above notation, that $\nu = \nu^V$ is a probability measure on $(M^V; \mathcal{B}_V)$. If $I \subseteq V$, let ν^I denote the pushforward measure on $(M^I; \mathcal{B}_I)$. So for $A \in \mathcal{B}_I$, we have $\nu^I(A) = \nu(\pi_I^{-1}(A))$, where $\pi_I : M^V \to M^I$ is the projection map.

Recall that if ν is a probability measure on a σ-algebra $\mathcal{B}$ of subsets of a set N, then $L^\infty(\mathcal{B})$ denotes the space of $\mathcal{B}$-measurable functions $N \to \mathbb{R}$ which are essentially bounded, that is, are bounded outside a set of measure 0.

Henceforth, we shall assume that the following conditions on ν hold.

- Definability: For all $J \subseteq V$ and $B \in \mathcal{B}_V$, the function $x_J \mapsto \nu^{V\setminus J}(B(x_J))$ is $\mathcal{B}_J$-measurable.
- Fubini: Suppose $J \subseteq V$ and $f \in L^\infty(\nu^V)$. Then $\int f \, d\nu^V = \iint f \, d\nu^J \, d\nu^{V\setminus J}$.

Remarks 2.2. (1) It would be more correct to refer to the definability condition as "Borel definability", but we will not do this.

(2) It suffices to check that the definability property holds for all $B \in \mathcal{B}_V^0$, as the set of elements of $\mathcal{B}_V$ for which it holds is a σ-subalgebra.

(3) The definability property is a weaker requirement than asking that ν be invariant (over the empty set, or a small submodel).

(4) The definability property implies that, in the statement of the Fubini condition, the map

$$\bar{x}_{V\setminus J} \mapsto \int f(\bar{x}_J \bar{x}_{V\setminus J}) \, d\nu^J(\bar{x}_J)$$

is $\mathcal{B}_{V\setminus J}$-measurable for almost all $\bar{x}_{V\setminus J} \in M^{V\setminus J}$. This is a standard argument using approximation by indicator functions of sets in $\mathcal{B}_V$. The same sort of argument shows that it suffices to check the Fubini condition in the case where f is the indicator function 1_B of a set $B \in \mathcal{B}_V^0$.

The following is Towsner's infinitary analogue of the hypergraph removal lemma. We refer to Towsner [2010; 2018] for a discussion of the origins of the proof and the finitary versions of this. The statement follows by combining Theorem 5.3 and Lemma 5.4 of [Towsner 2018]. Theorem 5.3 of [Towsner 2018] holds under weaker conditions than the Fubini property (involving the notion of J-*regularity* of ν^V), but we will not make use of this. Lemma 5.4 of [Towsner 2018] states that the definability and Fubini conditions imply J-regularity of ν^V for all $J \subseteq V$.

Theorem 2.3 [Towsner 2018, Theorem 5.3]. *Suppose $\mathcal{M}$ is sufficiently saturated and $\mathcal{B}_V^0$ consists of the definable subsets of M^V. Suppose ν^V is a probability measure on $\mathcal{B}^V$ which satisfies the definability and Fubini conditions. Let $k < n = |V|$ and $\mathcal{J} = [V]^k$, the set of k-subsets from V.*

Let $A_I \in \mathcal{B}_{V,I}$ for $I \in \mathcal{J}$. Suppose there is $\delta > 0$ such that whenever $B_I \in \mathcal{B}_{V,I}^0$ are such that $\nu^V(A_I \setminus B_I) < \delta$, then $\bigcap_{I\in\mathcal{J}} B_I \neq \varnothing$.

Then $\nu^V\big(\bigcap_{I\in\mathcal{J}} A_I\big) > 0$. □

2B. ***A model-theoretic corollary.*** In the following, we give model-theoretic conditions which allows us to verify the hypotheses in Theorem 2.3. The setup is:

- $\mathcal{M}$ is an $\aleph_1$-saturated structure in a countable language L.
- $V = \{1, \ldots, n\}$ is a set of indices (each associated to a particular sort); we let $J = \{1, \ldots, n-1\} \subseteq V$, and $\mathfrak{J}$ is the set of $(n-1)$-subsets of V.
- For each $I \subseteq V$, $\mathcal{B}_I^0$ is the Boolean algebra of M-definable subsets of M^I.
- $\nu = \nu^V$ is a probability measure on $\mathcal{B}_V$ which satisfies the definability and Fubini conditions.

For $I \subseteq V$ let π_I denote the projection map $M^V \to M^I$ and denote by ν^I the pushforward measure induced on $\mathcal{B}_I$ by ν. Each ν^I also satisfies the corresponding definability and Fubini properties.

Theorem 2.4. *With the above notation and assumptions, suppose $E \in \mathcal{B}_V$ is such that*

(a) $\nu^J(\pi_J(E)) > 0$;

(b) *there is $l \in \mathbb{N}$ such that for all $I \in \mathfrak{J}$ and $\bar{a} \in M^I$, we have that $\pi_I^{-1}(\bar{a}) \cap E$ has at most l elements*;

(c) *there is $k > 0$ such that if $F \in \mathcal{B}_V^0$, then $\nu^J(\pi_J(F \cap E)) \leq k\nu^I(\pi_I(F \cap E))$ for all $I \in \mathfrak{J}$.*

Then $\nu^V(\{\bar{b} \in M^V : \pi_I(\bar{b}) \in \pi_I(E)$ for all $I \in \mathfrak{J}\}) > 0$.

Remarks 2.5. We make some comments about the conditions on E. By the second condition, we should not expect that $\nu(E) > 0$. However, suppose that we also have a measure λ on the definable subsets of E with $\lambda(E) > 0$ and $r, s > 0$ such that for all $F \in \mathcal{B}_0^V$ and $I \neq J$ we have

$$r\nu^J(\pi_J(F \cap E)) \leq \lambda(F \cap E) \leq s\nu^I(\pi_I(F \cap E)).$$

Then $\nu^J(\pi(F \cap E)) \leq (s/r)\nu^I(\pi_I(F \cap E))$, so the third condition holds.

In general, without assuming the existence of such a λ, we can define a measure ν_I^J on $\pi_J(E)$ by setting $\nu_I^J(X) = \nu^I\big(\pi_I(\pi_J^{-1}(X) \cap E)\big)$. Condition (c) implies that ν^J is absolutely continuous with respect to ν_I^J and k is a bound on the Radon–Nikodým derivative.

Before proving Theorem 2.4 we note the following lemmas.

Lemma 2.6. *With the notation as in Theorem 2.4, suppose $F \subseteq E$ is a countable intersection of sets in $\mathcal{B}_V^0$ with E. Then*:

(1) $\nu^J(\pi_J(F)) \leq k\nu^I(\pi_I(F))$.

(2) *If $J \neq I \in \mathfrak{J}$ and $C \in \mathcal{B}_I^0$ then*

$$\nu^J\big(\pi_J(F \setminus \pi_I^{-1}(C))\big) \geq \nu^J(\pi_J(F)) - k\nu^I(C \cap \pi_I(F)).$$

(3) *If $J \neq I \in \mathfrak{J}$ and $B \in \mathcal{B}^0_I$, then*

$$\nu^J\big(\pi_J(F \cap \pi_I^{-1}(B))\big) \geq \nu^J(\pi_J(F)) - k\nu^I(\pi_I(F) \setminus B).$$

Proof. (1) Write $F = E \cap \bigcap_{i<\omega} F_i$, where each F_i is in $\mathcal{B}^0_V$. We can assume that $F_i \supseteq F_{i+1}$. Then $\aleph_1$-saturation implies $\pi_J(F) = \bigcap_{i<\omega} \pi_J(E \cap F_i)$ and $\nu^J(\pi_J(F)) = \inf\{\nu^J(\pi_J(E \cap F_i)) : i < \omega\}$. By assumption on E, we have $\nu^J(\pi_J(E \cap F_i)) \leq k\nu^I(\pi_I(E \cap F_i))$ for each i; taking the limit gives what we require.

(2) By (1) we have

$$\nu^J\big(\pi_J(\pi_I^{-1}(C) \cap F)\big) \leq k\nu^I(\pi_I(F) \cap C).$$

Of course, $\pi_J(F) = \pi_J(\pi_I^{-1}(C) \cap F) \cup \pi_J(F \setminus \pi_I^{-1}(C))$, so

$$\nu^J(\pi_J(F)) \leq \nu^J\big(\pi_J(\pi_I^{-1}(C) \cap F)\big) + \nu^J\big(\pi_J(F \setminus \pi_I^{-1}(C))\big).$$

Putting these together gives the required result.

(3) Apply (2), taking C to be the complement of B. □

Lemma 2.7. *Suppose $E \in \mathcal{B}_V$ with $\nu^J(\pi_J(E)) > 0$ and, for all $\bar{a} \in E$, we have that $\pi_J^{-1}(\pi_J(\bar{a})) \cap E$ has at most l elements. Let $X \subseteq M$ be a countable set over which E is definable.*

(1) *There is some $F \in \mathcal{B}_V(X)$ with $F \subseteq E$, a natural number r and an $L(X)$-formula $\psi(\bar{x})$ such that $\nu^J(\pi_J(F)) > 0$, and if $\bar{a} \in F$, then $\psi(\bar{a}_J, x_n)$ isolates $\mathrm{tp}^{\mathcal{M}}(a_n/\bar{a}_J X)$, and this type has precisely r solutions in $\mathcal{M}$. The set F can be taken to be a countable intersection of sets in $\mathcal{B}^0_V(X)$ with E.*

(2) *If X is chosen so that r in (1) is minimal, then for countable $Y \supseteq X$ and for almost all $\bar{a}_J \in \pi_J(F)$, if $(\bar{a}_J, a_n) \in F$, then $\psi(\bar{a}_J, x_n)$ isolates $\mathrm{tp}^{\mathcal{M}}(a_n/\bar{a}_J Y)$ (and therefore this type has the same solutions as $\mathrm{tp}^{\mathcal{M}}(a_n/\bar{a}_J X)$).*

Proof. (1) For each V-variable formula $\psi(\bar{x}) \in L(X)$ and $r \leq l$, consider the set $E_{\psi,r}$ consisting of those $(a_1, \ldots, a_n) \in E$ such that the formula $\psi(a_1, \ldots, a_{n-1}, x_n)$ isolates $\mathrm{tp}(a_n/a_1, \ldots, a_{n-1}, X)$ and this type has r solutions in $\mathcal{M}$. As E is defined over X, all of these solutions lie in E. Note that $E_{\psi,r}$ is defined by the conjunction of E, and

$$\bigwedge_{\chi \in L(X)} \Big(\psi(x_1, \ldots, x_n) \wedge (\exists^{=r} x_n)\psi(x_1, \ldots, x_n) \\ \wedge (\forall y)\big(\psi(x_1, \ldots, x_{n-1}, y) \to \big(\chi(x_1, \ldots, x_n) \leftrightarrow \chi(x_1, \ldots, x_{n-1}, y)\big)\big)\Big),$$

so is in $\mathcal{B}_V(X)$. Moreover, $\bigcup_{\psi;\, r \leq l} E_{\psi,r} = E$ (by the algebraicity). So as this is a countable union, there are ψ and $r \leq l$ with $\nu^J(\pi_J(E_{\psi,r})) > 0$. Then $F = E_{\psi,r}$ has the required properties.

(2) Let $Y \supseteq X$ be a countable subset of M and consider

$$E' = \{\bar{a} \in F : \psi(\bar{a}_J, x_n) \text{ does not isolate } \operatorname{tp}(a_n/\bar{a}_J Y)\}.$$

As in (1), we have $E' \in \mathcal{B}_V(Y)$. Suppose for a contradiction that $\nu^J(\pi_J(E')) > 0$. Applying (1) we obtain $F' \in \mathcal{B}_V(Y)$ with $F' \subseteq E'$ and $\nu^J(\pi_J(F')) > 0$, some $r' \in \mathbb{N}$ and an $L(Y)$-formula ψ' such that for all $\bar{a} \in F'$, $\psi'(\bar{a}_J, x_n)$ isolates $\operatorname{tp}(a_n/\bar{a}_J Y)$ and the latter has r' solutions. By definition of E' we have $r' < r$ and this contradicts the choice of r. Thus $\nu^J(\pi_J(E')) = 0$ and the result follows. □

We now prove Theorem 2.4.

Proof of Theorem 2.4. From Lemma 2.7(2), there is a countable subset X of M containing the parameters for E and a countable intersection F of X-definable sets with E such that

- $\nu^J(\pi_J(F)) > 0$;
- if $(a_1, \ldots, a_{n-1}, a_n), (a_1, \ldots, a_{n-1}, a'_n) \in F$, then $\operatorname{tp}^{\mathcal{M}}(a_n/a_1, \ldots, a_{n-1}, X) = \operatorname{tp}^{\mathcal{M}}(a'_n/a_1, \ldots, a_{n-1}, X)$;
- if $Y \supseteq X$ is countable, then for almost all $\bar{a} \in F$, the types $\operatorname{tp}^{\mathcal{M}}(a_n/\bar{a}_J X)$ and $\operatorname{tp}^{\mathcal{M}}(a_n/\bar{a}_J Y)$ have the same solutions.

To see the second point here, note that the two types are isolated by the same formula, so must be equal. The other points are directly from Lemma 2.7.

For $I \in \mathfrak{J}$, let $A_I = \pi_I^{-1}(\pi_I(F))$. So of course, $A_I \in \mathcal{B}_{V,I}$ and $F \subseteq \bigcap_{I \in \mathfrak{J}} A_I$. We verify that the hypotheses of Theorem 2.3 hold.

Let $\delta > 0$ (to be fixed later) and $B_I \in \mathcal{B}^0_{V,I}$ with $\nu^V(A_I \setminus B_I) < \delta$. Note that $\nu^V(A_I) = \nu^I(\pi_I(F))$ and similarly $\nu^I(\pi_I(F) \setminus \pi_I(B_I)) = \nu^V(A_I \setminus B_I)$. Therefore, with k as in condition (c) of Theorem 2.4 and $I \neq J$, Lemma 2.6(3) gives

$$\nu^J(\pi_J(F \cap B_I)) \geq \nu^J(\pi_J(F)) - k\nu^I(\pi_I(F) \setminus \pi_I(B_I)) > \nu^J(\pi_J(F)) - k\delta.$$

This also holds with $I = J$, as $k \geq 1$.

Now let $\eta = \nu^J(\pi_J(F))$ (so $\eta > 0$, by choice of F) and $\delta = \frac{1}{2}\eta k n$. We obtain, for all $I \in \mathfrak{J}$,

$$\nu^J(\pi_J(F \cap B_I)) \geq \left(1 - \tfrac{1}{2}n\right)\eta.$$

The measure of the union of the complements of the sets $\pi_J(F \cap B_I)$ in $\pi_J(F)$ is therefore at most $\frac{1}{2}\eta$, and so

$$\nu^J\left(\bigcap_{I \in \mathfrak{J}} \pi_J(F \cap B_I)\right) \geq \tfrac{1}{2}\eta.$$

Let Y be the union of X and the parameter sets of the B_I. Then we can find $\bar{b}_J = (b_1, \ldots, b_{n-1}) \in \bigcap_I \pi_J(F \cap B_I)$ such that if $(\bar{b}_J, b_n) \in F$ and $(\bar{b}_J, b'_n) \in F$, then they have the same type over Y. Indeed, almost all $\bar{b}_J \in \pi_J(F)$ have this property, by our conditions on F.

Take $b_n \in M$ with $\bar{b} = (b_1, \dots, b_{n-1}, b_n) \in F$. We show that $(b_1, \dots, b_n) \in \bigcap_I B_I$, and thus the hypotheses of Theorem 2.3 hold.

Clearly $\bar{b} \in B_J$. Take $I \neq J$. Because $(b_1, \dots, b_{n-1}) \in \pi_J(F \cap B_I)$, there is $(b'_1, b'_2, \dots, b'_n) \in F \cap B_I$ such that $(b'_1, \dots, b'_{n-1}) = (b_1, \dots, b_{n-1})$. Therefore $(b_1, \dots, b_{n-1}, b_n), (b_1, \dots, b_{n-1}, b'_n) \in F$, and thus b_n and b'_n have the same type over $Y \cup \{b_1, \dots, b_{n-1}\}$. As B_I is defined over Y and $(b_1, \dots, b_{n-1}, b'_n) \in B_I$, it follows that $(b_1, \dots, b_{n-1}, b_n) \in B_I$, as required.

We have shown $\bigcap_I B_I \neq \emptyset$, so Theorem 2.3 applies to give that $\nu\big(\bigcap_I A_I\big) > 0$. As $\bigcap_I A_I \subseteq \{\bar{b} \in \mathcal{M}^V : \pi_I(\bar{b}) \in \pi_I(E) \text{ for all } I \in \mathcal{J}\}$, we have the result. □

3. Examples and applications

3A. ***Pseudofinite structures and Szemerédi's theorem.*** In [Towsner 2010] (and [Towsner 2018, Section 5]), the structure $\mathcal{M}$ is an ultraproduct of finite structures $(F_i : i < \omega)$ and the measures arise by taking the standard part of ultraproducts of normalised counting measures on the F_i. The original language is enriched to ensure definability of the measure. The Fubini property then follows, as we are dealing with counting measures.

In [Towsner 2010, Section 2], Szemerédi's theorem is deduced from Theorem 2.3 in the following way (we do not give the details: the point is to explain where the statement of Theorem 2.4 comes from). The original language is that of abelian groups (written additively) and there is a predicate $A(\cdot)$ for a subset of the group. Each F_i is cyclic of prime order (increasing with i) and $A[F_i]$ is some subset of F_i. Denoting the ultraproduct (in the enriched language) by G, the main assumption is that the measure of $A[G]$ is strictly positive.

So $(G, +)$ is a torsion-free, divisible abelian group, and if $n \in \mathbb{N}$ and $n \geq 3$, we have a definable measure ν^n on the definable subsets of G^n which satisfies the hypotheses of Theorem 2.4. The measure is invariant under definable bijections (in particular, under translations and taking i-th roots). Let

$$E = \left\{ \left(x_1, \dots, x_{n-1}, \sum_{i<n} x_i \right) : \sum_{i<n} i x_i \in A \right\}.$$

This is definable, and in the notation of Theorem 2.4, $\pi_J(E) = \nu^1(A) > 0$ (using the divisibility of G and invariance of the measure under definable bijections). The projection maps π_I (with $|I| = n-1$) are injective on E and thus the remaining two conditions in Theorem 2.4 hold (with $k = l = 1$).

So, by Theorem 2.4, there is some $\bar{b} = (b_1, \dots, b_n) \in G^n$ such that $\pi_I(\bar{b}) \in \pi_I(E)$ for all I of size $n-1$, and by positivity of the measure, we can take $d = b_n - \sum_{i<n} b_j$

to be nonzero. The definition of E means that if we set $a = \sum_{i<n} i b_i$, then $a, a+d, \dots, a+(n-1)d \in A$. Therefore, as $d \neq 0$, we have an n-term arithmetic progression in A.

3B. ***An amalgamation result in MS-measurable structures.*** The notion of a *measurable structure* was introduced by Macpherson and Steinhorn [2008], following on from observations of Chatzidakis, van den Dries and Macintyre in [Chatzidakis et al. 1992]. Elwes and Macpherson [2008] give a survey of results and open questions. Following [Kestner and Pillay 2011], we refer to this notion as *MS-measurability*.

We recall the definition of MS-measurability [Macpherson and Steinhorn 2008, Definition 5.1]. For a (first-order) L-structure $\mathcal{M}$ we denote by $\mathrm{Def}(\mathcal{M})$ the collection of all nonempty parameter definable subsets of M^n (for all $n \geq 1$).

Definition 3.1. A structure $\mathcal{M}$ is *MS-measurable* if there is a *dimension–measure* function $h : \mathrm{Def}(\mathcal{M}) \to \mathbb{N} \times \mathbb{R}^{>0}$ satisfying the following, where we write $h(X) = (\dim(X), \mu(X))$:

(i) If X is finite (and nonempty) then $h(X) = (0, |X|)$.

(ii) For every formula $\phi(\bar{x}, \bar{y})$ there is a finite set $D_\phi \subseteq \mathbb{N} \times \mathbb{R}^{>0}$ of possible values for $h(\phi(\bar{x}, \bar{a}))$ (with $\bar{a} \in M^n$), and for each such value, the set of $\bar{a}$ giving this value is 0-definable.

(iii) Fubini property: Suppose $X, Y \in \mathrm{Def}(\mathcal{M})$ and $f : X \to Y$ is a definable surjection. By (ii), Y can be partitioned into disjoint definable sets $Y_1, \dots, Y_r$ such that $h(f^{-1}(y))$ is constant, equal to (d_i, m_i), for $y \in Y_i$. Let $h(Y_i) = (e_i, n_i)$. Let c be the maximum of $d_i + e_i$ and suppose this is attained for $i = 1, \dots, s$. Then $h(X) = (c, m_1 n_1 + \cdots + m_s n_s)$.

In the above, $\dim(X)$ is the *dimension* and $\mu(X)$ the *measure* of X. Clearly we can normalise and assume that $\mu(M) = 1$. We also extend the definition so that $\mu(\varnothing) = 0$. Note that MS-measurability is a property of the theory of M, so any elementary extension or submodel of $\mathcal{M}$ is MS-measurable if $\mathcal{M}$ is. As observed in [Macpherson and Steinhorn 2008, Remark 5.2], measurability implies supersimplicity and dimension dominates D-rank, but is not necessarily equal to it. By [Macpherson and Steinhorn 2008, Proposition 5.10], the dimension–measure function extends to definable subsets of $\mathcal{M}^{\mathrm{eq}}$.

We suppose (for convenience) that L is countable and suppose that $\mathcal{M}$ is an $\aleph_1$-saturated MS-measurable structure with dimension–measure function $h = (\dim, \mu)$. Let $S \in \mathrm{Def}(\mathcal{M})$ be infinite and let $\mathcal{B}^0_S$ denote the set of definable subsets of S. For $D \in \mathcal{B}^0_S$ we define

$$\nu^S(D) = \begin{cases} \mu(D)/\mu(S) & \text{if } \dim(D) = \dim(S), \\ 0 & \text{otherwise.} \end{cases}$$

If $X_1, X_2 \in \mathcal{B}_S^0$ are disjoint, then (iii) of Definition 3.1 (with Y a two-point set) shows that $\nu^S(X_1 \cup X_2) = \nu^S(X_1) + \nu^S(X_2)$. So ν^S is a finitely additive probability measure on $\mathcal{B}_S^0$ and it therefore extends uniquely to a probability measure on $\mathcal{B}_S$, which we will also denote by ν^S.

Now suppose that $S_1, \ldots, S_n \in \mathrm{Def}(\mathcal{M})$ are infinite and $S = S_1 \times \cdots \times S_n$. If $I \subseteq V = \{1, \ldots, n\}$, let S_I be the product of the S_i for $i \in I$. As previously, $\pi_I : S \to S_I$ is the projection map. By considering this, (iii) in Definition 3.1 gives that $\dim(S) = \dim(S_I) + \dim(S_{V \setminus I})$ and $\mu(S) = \mu(S_I)\mu(S_{V \setminus I})$.

Let $\nu = \nu^V = \nu^S$. If $I \subseteq V$, then the pushforward measure ν^I on $\mathcal{B}_{S_I}$ obtained from ν and π_I is equal to ν^{S_I}, as defined above. Indeed, it suffices to check this for $D \in \mathcal{B}_{S_I}^0$. If $\dim(D) = \dim(S_I)$, then

$$\begin{aligned} \nu^I(D) &= \nu^V(D \times S_{V \setminus I}) \\ &= \mu(D \times S_{V \setminus I})/\mu(S) = \mu(D)\mu(S_{V \setminus I})/\mu(S) = \mu(D)/\mu(S_I), \end{aligned}$$

and this is equal to $\nu^{S_I}(D)$. If $\dim(D) < \dim(S_I)$, then $\dim(D \times S_{V \setminus I}) < \dim(S)$, so both $\nu^I(D)$ and $\nu^{S_I}(D)$ are zero.

The definability and Fubini properties given in Section 2A hold for the ν^I, using (ii) and (iii) of Definition 3.1 (see Remarks 2.2).

From Theorem 2.4 we obtain the following, which can be seen as a weak form of independent n-amalgamation:

Corollary 3.2. *Suppose $\mathcal{M}$ is an MS-measurable structure and $S_1, \ldots, S_n \in \mathrm{Def}(\mathcal{M})$ are infinite and defined over a finite set $C \subset M$. Let $S = S_1 \times \cdots \times S_n$ and suppose $E \subseteq S$ is a C-definable subset such that*

(a) $\dim(\pi_I(E)) = \sum_{i \in I} \dim(S_i)$ *for all* $I \in [n]^{n-1}$, *and*

(b) *if* $(b_1, \ldots, b_n) \in E$, *then* $b_i \in \mathrm{acl}(C \cup \{b_j : j \neq i\})$.

Then

$$\dim\bigl\{\bar{b} \in S : \pi_I(\bar{b}) \in \pi_I(E) \textit{ for all } I \in [n]^{n-1}\bigr\} = \dim(S).$$

Remarks 3.3. Assumptions (a) and (b) in Corollary 3.2 imply that the S_i have the same dimension. Indeed, $\sum_{j<n} \dim(S_j) = \dim(\pi_J(E)) = \dim(E) = \dim(\pi_I(E)) = \sum_{i \in I} \dim(S_i)$ for all $I \in [n]^{n-1}$. So $\dim(S_j) = \dim(S_n)$ for all $j < n$.

We now prove Corollary 3.2.

Proof. We may assume that $\mathcal{M}$ is $\aleph_1$-saturated. We check that the three conditions of Theorem 2.4 hold.

By (a), $\dim(\pi_J(E)) = \dim(S_J)$, so $\nu^J(\pi_J(E)) = \mu(\pi_J(E))/\mu(S_J) > 0$.

As E is definable, by compactness we have a uniform bound l on the algebraicity in assumption (b). This gives the second condition required by Theorem 2.4.

Suppose $I \in [n]^{n-1}$. The restriction of the projection map $E \to \pi_I(E)$ has finite fibres, of size at most l. Suppose $X \subseteq E$ is definable. If we decompose $\pi_I(X)$ according to the size of the fibres $X \to \pi_I(X)$ and apply (i) and (iii) of Definition 3.1, we obtain

$$\mu(\pi_I(X)) \leq \mu(X) \leq l\mu(\pi_I(X)).$$

Thus

$$\mu(\pi_J(X)) \leq \mu(X) \leq l\mu(\pi_I(X)).$$

If $\dim(X) = \dim(E)$, then $\dim(\pi_I(X)) = \dim(\pi_I(E)) = \dim(S_I)$ (by (a)) and we obtain

$$\nu^J(\pi_J(X)) \leq l\frac{\mu(S_I)}{\mu(S_J)}\nu^I(\pi_I(X)).$$

If $\dim(X) < \dim(E)$ then the inequality is also true, as both sides are zero. So we have the third condition required by Theorem 2.4.

So, by Theorem 2.4,

$$\nu^V(\{\bar{b} \in S : \pi_I(\bar{b}) \in \pi_I(E) \text{ for all } I \in [n]^{n-1}\}) > 0,$$

and the conclusion follows. □

3C. ***Further examples.*** If Th($\mathcal{M}$) is NIP, then *generically stable* measures (see [Hrushovski and Pillay 2011] or [Simon 2015]) provide examples of measures satisfying the definability and Fubini conditions. More precisely, suppose $\nu_{x_1}, \ldots, \nu_{x_n}$ are generically stable measures for $\mathcal{M}$ (in the indicated variables) and let $\nu^V = \nu_{x_1} \otimes \cdots \otimes \nu_{x_n}$. Then ν^V has the definability and Fubini properties, and therefore Theorems 2.3 and 2.4 hold. It would be interesting to know whether either of these results is saying something new, or at least nontrivial, in this context.

4. MS-measurability and the Hrushovski construction

In [Elwes and Macpherson 2008, Definition 3.13], a complete theory is defined to be *unimodular* if in any model $\mathcal{M}$, whenever $f_i : X \to Y$ are definable k_i-to-1 surjections in $\mathcal{M}^{\text{eq}}$ (for $i = 1, 2$), then $k_1 = k_2$. (See [Kestner and Pillay 2011] for comments on this and, in particular, on why it should more properly be termed weak unimodularlity.) An MS-measurable structure is necessarily superstable of finite SU-rank and unimodular, and Question 7 of [Elwes and Macpherson 2008] asks whether the converse holds. Unimodularity is implied by ω-categoricity [Elwes and Macpherson 2008, Proposition 3.16], and in a similar vein, Question 2 of [Elwes and Macpherson 2008] asks whether a MS-measurable ω-categorical structure is necessarily one-based. For both of these questions the key examples to be considered are Hrushovski's non-locally-modular supersimple ω-categorical structures [1997; 1988]. In this section we apply Corollary 3.2 to show that *some* of Hrushovski's examples are *not* MS-measurable. In particular, this answers Question 7 of [Elwes and Macpherson 2008]: there is a supersimple, finite rank unimodular theory (even, ω-categorical, SU-rank 1) which is not MS-measurable.

4A. ***The Hrushovski construction for ω-categorical structures.*** We recall briefly some details of the construction method. The original version is in [Hrushovski 1988], where it is used to provide a counterexample to Lachlan's conjecture, and in [Hrushovski 1997], where it is used to construct a nonmodular, supersimple $\aleph_0$-categorical structure. The book [Wagner 2000] is a very convenient reference for this (see Section 6.2.1). Generalisations and reworkings of the method (particularly relating to simple theories) are also to be found in [Evans 2002]. We will restrict to the simplest form of the construction appropriate for producing ω-categorical structures of SU-rank 1.

We work with a finite relational language $L = \{R_i : i \leq m\}$. For later use, it will be convenient to assume that this contains some 3-ary relation R. Recall that if B and C are L-structures with a common substructure A then the *free amalgam* $B \amalg_A C$ of B and C over A is the L-structure whose domain is the disjoint union of B and C over A and whose atomic relations are precisely those of B together with those of C. Let $\overline{\mathcal{K}}$ be the class of L-structures and denote by $\mathcal{K}$ the finite structures in $\overline{\mathcal{K}}$.

For $A \in \mathcal{K}$ define the *predimension* $\delta(A)$ to be equal to $|A| - \sum_i |R_i[A]|$. If $A \subseteq B \in \mathcal{K}$ write $A \leq B$ to mean $\delta(A) < \delta(B')$ for all $A \subset B' \subseteq B$. (We sometimes say that A is *self-sufficient* in B.) For structures in $\mathcal{K}$, one has

- if $X \subseteq B$ and $A \leq B$, then $X \cap A \leq X$;
- if $A \leq B \leq C$, then $A \leq C$.

Consequently (see [Wagner 2000, Corollary 6.2.8]), for each $B \in \mathcal{K}$ there is a closure operation given by $\mathrm{cl}_B(X) = \bigcap\{A : X \subseteq A \leq B\} \leq B$ for $X \subseteq B$. Of course, if $B \leq C \in \mathcal{K}$ and $X \subseteq B$, then $\mathrm{cl}_B(X) = \mathrm{cl}_C(X)$.

The relation $\leq$ can be extended to infinite structures so that the above properties still hold: if $M \in \overline{\mathcal{K}}$ and $A \subseteq M$, write $A \leq M$ to mean that $A \cap X \leq X$ for all finite $X \subseteq M$.

If $A, B \in \overline{\mathcal{K}}$, an embedding $\alpha : A \to B$ with $\alpha(A) \leq B$ is called a $\leq$-embedding.

Now consider $\overline{\mathcal{K}}_0$, the class of $B \in \overline{\mathcal{K}}$ with $\varnothing \leq B$. Equivalently, if $A \subseteq B$ is finite and nonempty, then $\delta(A) > 0$. Let $\mathcal{K}_0$ be the finite structures in $\overline{\mathcal{K}}_0$. Any structure B in $\overline{\mathcal{K}}_0$ carries a notion of *dimension* d^B associated to the predimension δ and a notion of d^B*-independence*. If $X, Y \subseteq B$ are finite, write $d^B(X) = \delta(\mathrm{cl}_B(X)) = \min\{\delta(Y) : X \subseteq Y \subseteq B\}$ and $d^B(X/Y) = d^B(X \cup Y) - d^B(Y)$. If the ambient structure B is clear from the context, then we omit it from the notation. Say that finite X and Z are *d-independent* over Y (in B) if $d^B(X/YZ) = d^B(X/Y)$. In particular, this implies $\mathrm{cl}_B(XY) \cap \mathrm{cl}_B(YZ) = \mathrm{cl}_B(Y)$. (Here, we use the usual shorthand of YZ for $Y \cup Z$.) For the particular predimension which we have given, it can be shown that cl_B satisfies the exchange condition, and therefore gives a pregeometry; furthermore, d^B is the dimension in this pregeometry.

We look at a version of the construction (also from [Hrushovski 1997]) where closure is uniformly locally finite. For this, we have a continuous, increasing $f : \mathbb{R}^{\geq 0} \to \mathbb{R}$ with $f(x) \to \infty$ as $x \to \infty$ and we consider $\mathcal{K}_f = \{A \in \mathcal{K}_0 : \delta(X) \geq f(|X|)$ for all $X \subseteq A\}$. For suitable choice of f (call these *good* f), $(\mathcal{K}_f, \leq)$ has the free $\leq$-amalgamation property: if $A_0 \leq A_1, A_2 \in \mathcal{K}_f$, then $A_i \leq A_1 \amalg_{A_0} A_2 \in \mathcal{K}_f$. In this case we have an associated *generic structure* M_f (see [Wagner 2000, Theorem 6.2.13]). This is a countable structure characterised by the following properties:

(i) M_f is the union of a chain of finite self-sufficient substructures, all in $\mathcal{K}_f$.

(ii) $\leq$-extension property: If $A \leq M_f$ is finite and $A \leq B \in \mathcal{K}_f$, then there is a $\leq$-embedding $\beta : B \to M_f$ with $\beta(a) = a$ for all $a \in A$.

Equivalently, $\mathcal{K}_f$ is the class of finite substructures of M_f, and isomorphisms between finite self-sufficient substructures of M_f extend to automorphisms of M_f (we refer to the latter property as $\leq$-homogeneity). Because of the function f, closure in M_f is uniformly locally finite and (using free amalgamation and the $\leq$-extension property) it is equal to algebraic closure [Wagner 2000, Lemma 6.2.17]. It then follows from $\leq$-homogeneity that M_f is ω-categorical and the type of a tuple is determined by the isomorphism type of its closure.

Remarks 4.1. To construct good functions, we can take f which are piecewise smooth, and where the right derivative f' satisfies $f'(x) \leq 1/(x+1)$ and is nonincreasing. The latter condition implies that $f(x+y) \leq f(x) + yf'(x)$ (for $y \geq 0$). It can be shown that under these conditions, $\mathcal{K}_f$ has the free $\leq$-amalgamation property. (This is originally from [Hrushovski 1988]; see also [Wagner 2000, Example 6.2.27] or [Evans 2002, Lemma 3.3].)

Remarks 4.2 ([Hrushovski 1997]; see also [Wagner 2000, Example 6.2.27; Evans 2002, Corollary 2.24, Theorem 3.6]). If f also satisfies the slower growth condition

$$f(3x) \leq f(x) + 1,$$

then the structure M_f is supersimple of SU-rank 1. Moreover, for tuples $\bar{a}$ and $\bar{b}$ in M_f, we have $\mathrm{SU}(\mathrm{tp}(\bar{b}/\bar{a})) = d(\bar{b}/\bar{a})$. To see the latter, note that (by additivity of both sides) it suffices to prove this when $\bar{b}$ is a single element b. Now, $d(b/\bar{a})$ is a natural number and at most $\delta(b)$, so is 0 or 1. If it is 0, then $b \in \mathrm{acl}(\bar{a})$ so $\mathrm{SU}(b/\bar{a}) = 0$. Thus, it suffices to show that if $\mathrm{tp}(b/\bar{a})$ divides over $\varnothing$, then $d(b/\bar{a}) < d(b/\varnothing)$. This is done (in greater generality) in the above references.

4B. ***The dimension function.*** For the rest of the section suppose that f is a good function as in Remarks 4.1 and M_f is the corresponding generic structure. We suppose that $h = (\dim, \mu) : \mathrm{Def}(M_f) \to \mathbb{R}^{>0}$ is a dimension–measure function. In this subsection we relate dim to the dimension d coming from the predimension (which will be the same as SU-rank if M_f is simple), and the measure will not be used.

Notation 4.3. For tuples $\bar{a}$ and $\bar{b}$ in M_f, let $\mathrm{loc}(\bar{b}/\bar{a})$, the *locus of $\bar{b}$ over $\bar{a}$*, be the set of realisations in M_f of $\mathrm{tp}_L(\bar{b}/\bar{a})$, the L-type of $\bar{b}$ over $\bar{a}$. By ω-categoricity, this is definable by an L-formula with parameters from $\bar{a}$. Let $\dim(\bar{b}/\bar{a})$ denote the dimension of this set.

The Fubini property in MS-measurability implies that dim is additive: $\dim(\bar{b}/\bar{a}) = \dim(\bar{a}\bar{b}/\varnothing) - \dim(\bar{a}/\varnothing)$. We also have $\dim(M_f^n) = n \dim(M_f)$. Note the existence of dim-generic points: if $D \in \mathrm{Def}(M_f)$ is definable over a finite tuple $\bar{a}$, then $\dim(D) = \max\{\dim(\bar{b}/\bar{a}) : \bar{b} \in D\}$. From this we deduce that if $D' \subseteq D$ is definable, then $\dim(D') \leq \dim(D)$. A further property of dim which we require is the weak algebraicity property that if $\bar{b} \in \mathrm{acl}(\bar{a})$, then $\dim(\bar{b}/\bar{a}) = 0$. Of course, d also has these properties.

Under these assumptions on dim (and the given conditions on f) we will show that dim is just a scaled version of the dimension d.

Theorem 4.4. *Suppose $f'(x) \leq \frac{1}{2}(1/(x+1))$. If $\bar{a}$ and $\bar{b}$ are finite tuples in M_f, then we have*

$$\dim(\bar{b}/\bar{a}) = \dim(M_f) d(\bar{b}/\bar{a}).$$

The theorem follows from the following (always assuming the given condition on f).

Proposition 4.5. *Let $\bar{a}, b \in M_f$ with $b \notin \mathrm{acl}(\bar{a})$ and $P = \mathrm{loc}(b/\bar{a})$. Then, for every $r \in \mathbb{N}$ and $\bar{y} \in M^r$, there is some $\bar{x} \in P^{r+2}$ with $\bar{y} \in \mathrm{acl}(\bar{x}\bar{a})$.*

We note that Marimon (unpublished work) shows that Theorem 4.4 holds for a wider class of Hrushovski constructions than we give here.

First we show how Theorem 4.4 follows from the proposition.

Proof of Theorem 4.4. By the additivity property of both dim and d, it will suffice to prove the statement when $\bar{b} = b$ is a single element. If $b \in \mathrm{acl}(\bar{a})$, then the statement holds as both sides of the equation are zero, by the weak algebraicity property of dim and d. So now suppose that $b \notin \mathrm{acl}(\bar{a})$. Let $P = \mathrm{loc}(b/\bar{a})$, as in Proposition 4.5. Consider

$$Y = \{\bar{y} = (y_1, \ldots, y_r) \in M_f^r : y_1, \ldots, y_r \in \mathrm{acl}(\bar{x}\bar{a}) \text{ for some } \bar{x} \in P^{2+r}\}.$$

By ω-categoricity, this set is definable by an L-formula with parameters from $\bar{a}$ (for example, it is invariant under automorphisms of M_f fixing $\bar{a}$). Thus (by existence of generic points for dim) there is $\bar{c} \in Y$ with $\dim(Y) = \dim(\bar{c}/\bar{a})$. By definition of Y, there are $b_1, \ldots, b_{r+2} \in P$ with $\bar{c} \in \operatorname{acl}(\bar{a}b_1 \cdots b_{r+2})$. It follows (using weak algebraicity) that

$$\dim(Y) = \dim(\bar{c}/\bar{a}) \leq \dim(b_1 \cdots b_{r+2}/\bar{a}) \leq \dim(M_f^{r+2}) = (r+2)\dim(M_f).$$

But, by Proposition 4.5, we have $Y = M_f^r$. So

$$r\dim(M_f) = \dim(M_f^r) = \dim(Y) \leq (r+2)\dim(M_f).$$

Dividing by $(r+2)$ and letting $r \to \infty$, we obtain that $\dim(b/\bar{a}) = \dim(M_f)$. As $d(b/\bar{a}) = 1$, this gives $\dim(b/\bar{a}) = \dim(M_f)d(b/\bar{a})$, as required. □

The proof of Proposition 4.5 is a technical argument with Hrushovski constructions, so we relegate it to a separate section (Section 4D). Marimon's approach [2022a; 2022b] to proving non-MS-measurability of other examples of ω-categorical Hrushovski constructions avoids the need for a result such as Theorem 4.4.

Remarks 4.6. It is an open problem to determine whether any of the M_f are (or are not) pseudofinite. We note that Theorem 4.4 provides some information relevant to this question. Suppose that f is a good function with $f'(x) \leq \frac{1}{2}(1/(x+1))$ and $\mathcal{K}_f$ is the corresponding amalgamation class with generic structure M_f. *Assume* that M_f is elementarily equivalent to an ultraproduct $\mathcal{M} = \prod_U F_i$ of finite structures. Following [Hrushovski 2013], if $\Phi(\bar{x})$ is a formula with parameters from M, then the *coarse pseudofinite dimension* $\Delta(\Phi(\bar{x}))$ is the standard part of the nonstandard real $\prod_U \log|\Phi(F_i)|/\log|F_i|$. We will show that for every L-formula $\Phi(\bar{x})$ (without parameters), we have $\Delta(\Phi(\bar{x})) = d(\Phi(\bar{x}))$.

In principle, we could deduce the result from Theorem 4.4 as Δ has the properties required in the proof of Theorem 4.4, as long as we expand the language by dimension quantifiers so that it becomes continuous (see [Hrushovski 2013, Section 2.7]). However, it seems clearer to give a fuller argument which is essentially a modification of that given for Theorem 4.4.

If $\bar{a}$ is a finite tuple in M_f, let $\Phi_{\bar{a}}(\bar{x})$ denote an L-formula isolating $\operatorname{tp}(\bar{a}/\varnothing)$ (the L-type of $\bar{a}$ in M_f). Such a formula exists, by ω-categoricity. If $\bar{b}$ is another tuple, then $\Phi_{\bar{a}\bar{b}}(\bar{a}, \bar{y})$ isolates $\operatorname{tp}(\bar{b}/\bar{a})$.

Claim. *Suppose $\bar{a}$ is a k-tuple in M_f and $b \in M_f$. Suppose $\bar{u}$ is a k-tuple in $\mathcal{M}$ and $\mathcal{M} \models \Phi_{\bar{a}}(\bar{u})$. Then $\Delta(\Phi_{\bar{a}b}(\bar{u}, y)) = d(b/\bar{a})$.*

Proof of claim. If $d(b/\bar{a}) = 0$ then b is algebraic over $\bar{a}$. The size of $\operatorname{acl}(\bar{a})$ is bounded uniformly (actually, in k), so $\Phi_{\bar{a}b}(\bar{u}, y)$ has finitely many solutions in $\mathcal{M}$. Thus its pseudofinite dimension is 0.

Now suppose that $b \notin \operatorname{acl}(\bar{a})$, so that $d(b/\bar{a}) = 1$. Let $r \in \mathbb{N}$. There is a formula $C_r(y, x_1 \cdots x_{r+2}\bar{z})$ such that if $\Phi_{\bar{a}b}(\bar{a}'b_i)$ (for $i \leq r+2$), then $C_r(M_f, b_1 \cdots b_{r+2}, \bar{a}')$ is $\operatorname{acl}(b_1 \cdots b_{r+2}, \bar{a}')$. Let $K(r)$ bound the size of this algebraic closure.

The set Y in the proof of Theorem 4.4 is defined by $Y(\bar{y}, \bar{a})$, where $Y(\bar{y}, \bar{z})$ is the formula

$$\exists x_1, \ldots, x_{r+2} \bigwedge_{i \le r+2} \Phi_{\bar{a}b}(\bar{z}x_i) \wedge \bigwedge_{j \le r} C_r(y_j, x_1 \cdots x_{r+2}\bar{z}).$$

Moreover, by Proposition 4.5,

$$M_f \models (\forall \bar{z})\big(\Phi_{\bar{a}}(\bar{z}) \to \forall y_1, \ldots, y_r\, Y(y_1 \cdots y_r \bar{z})\big),$$

so this formula also holds in $\mathcal{M}$.

Suppose $\bar{u} \in \mathcal{M}$ and $\mathcal{M} \models \Phi_{\bar{a}}(\bar{u})$. Denote by $\bar{u}_i$ the k-tuple of i-th coordinates (in F_i) in $\bar{u}$. From the definition of Y, for almost all i, we have

$$|Y(F_i, \bar{u}_i)| \le K(r)^r |\Phi_{\bar{a}b}(\bar{u}_i, F_i)|^{r+2}.$$

Thus, as $Y(\mathcal{M}, \bar{u}) = M^r$, for almost all i,

$$K(r)^r |\Phi_{\bar{a}b}(\bar{u}_i, F_i)|^{r+2} \ge |F_i|^r.$$

As $|F_i| \to \infty$, we obtain $\Delta(\Phi_{\bar{a}b}(\bar{u}, y)) \ge r/(r+2)$. But r here is arbitrary, and thus $\Delta(\Phi_{\bar{a}b}(\bar{u}, y)) \ge 1$. The reverse inequality is trivial, so we have the claim. □

Now suppose that $\bar{b} = (b_1, \ldots, b_n)$ is an n-tuple in M_f. We show that if $\bar{u}$ is a tuple in $\mathcal{M}$ and $\mathcal{M} \models \Phi_{\bar{a}}(\bar{u})$, then $\Delta(\Phi_{\bar{a}\bar{b}}(\bar{u}, \bar{y})) = d(\bar{b}/\bar{a})$. The required formula for general L-definable sets follows as each such is a finite union of pairwise disjoint sets of this form. We may assume that $d(\bar{b}/\bar{a}) = n$ and we prove the result by induction on n. Let D be $\Phi_{\bar{a}\bar{b}}(\bar{u}, \mathcal{M}) \subseteq \mathcal{M}^n$ and $E = \Phi_{\bar{a}b_1 \cdots b_{n-1}}(\bar{u}, \mathcal{M})$. Let $f : D \to E$ be projection onto the first $n-1$ coordinates. By the claim, the fibres of f have coarse pseudofinite dimension $d(b_n/b_1 \cdots b_{n-1}\bar{a}) = 1$. By the induction hypothesis, $\Delta(E) = n-1$. Thus, by Lemma 2.8(4) of [Hrushovski 2013], $\Delta(D) = n-1+1 = n$, as required. (In order to apply the results from [Hrushovski 2013], we need to first enrich the language so that Δ becomes continuous, but this has no effect on the dimension of formulas in the original language.)

4C. *A structure which is not MS-measurable.*

Theorem 4.7. *There is an ω-categorical, supersimple structure M_f of* SU-*rank* 1 *which does not satisfy the amalgamation property in Corollary 3.2. In particular, M_f is not MS-measurable.*

Proof. We choose f so that $\mathcal{K}_f$ is a free amalgamation class; the generic M_f is supersimple of SU-rank 1; the independent amalgamation property, Corollary 3.2, does not hold. We are only interested in providing an example, so we choose economy of effort over elegance.

Take L to have a 3-ary relation R, a 10-ary relation S and a 11-ary relation U. Let $f(x) = \log_8(x+1)$. Then $f'(x) = \frac{1}{\ln 8}(1/(x+1)) < \frac{1}{2}(1/(x+1))$, and therefore, by Remarks 4.1, $\mathcal{K}_f$ is a free amalgamation class and the hypothesis on f in Theorem 4.4 holds. We also have $f(3x) \le f(x) + 1$, so by Remarks 4.2, the generic M_f is supersimple, with d-independence being the same as nonforking, and M_f is of SU-rank 1.

Consider the L-structure A with points $a_1, \dots, a_{10}, u_1, \dots, u_r$, where $r = 8^9 - 11$, and relations $S(a_1, \dots, a_{10})$ and $U(a_1, \dots, a_{10}, u_i)$ (for $i \le r$). Then $\delta(A) = 9$ and $|A| = 8^9 - 1$, so $\delta(A) \ge f(|A|)$. It is easy to check that for any $X \subset A$ we have $\delta(X) \ge f(|X|)$, so $A \in \mathcal{K}_f$. Moreover (in the notation of Corollary 3.2), for each $I \in [10]^9$, the tuple $\bar{a}_I$ is d-independent (in A) and has closure A. Note also that if $I \in [10]^8$, then $\bar{a}_I \le A$.

Suppose, for a contradiction, that the conclusion of Corollary 3.2 holds, where dim is given by SU-rank (in this case, given by the dimension function d). We will apply this where $n = 10$, $S = M_f^{10}$ and

$$E = \{\alpha(a_1 \cdots a_{10}) \mid \alpha : A \to M_f \text{ is an } \le\text{-embedding}\}.$$

Note that E is $\varnothing$-definable, the algebraic closure (equal to the $\le$-closure) of every element of E is isomorphic to A, and (by the $\le$-homogeneity of M_f) all elements of E have the same type over $\varnothing$.

Therefore, if the conclusion of Corollary 3.2 holds, there exists a d-independent set $B_0 = \{b_1, \dots, b_{10}\}$ of distinct elements of M_f with the property that for each $I \in [10]^9$ we have $\mathrm{acl}_{M_f}(\mathcal{B}_I) \cong A$ (via an isomorphism taking $\bar{b}_i \mapsto \bar{a}_i$), where $B_I = \{b_i : i \in I\}$. Let $B = \mathrm{acl}(B_0)$. By the d-independence, $\delta(B) = 10$ and we have $\mathrm{acl}(B_I) \cap \mathrm{acl}(B_{I'}) = B_I \cap B_{I'} = B_{I \cap I'}$ for $I \ne I' \in [10]^9$.

Thus

$$\begin{aligned} |B| &\ge |B_0| + \left| \bigcup_{I \in [10]^9} \mathrm{acl}(B_I) \setminus B_0 \right| \\ &= |B_0| + \sum_{I \in [10]^9} |\mathrm{acl}(B_I) \setminus B_0| \\ &\ge 10 + 10(8^9 - 1 - 9) = 10.8^9 - 90. \end{aligned}$$

So

$$f(|B|) \ge \log_8(10.8^9 - 89) > 10 = \delta(B),$$

and thus $B \notin \mathcal{K}_f$, a contradiction. So the amalgamation property in the conclusion of Corollary 3.2 does not hold, and in particular, M_f is not MS-measurable. □

4D. ***Proof of Proposition 4.5.*** Before the proof, we give a technical lemma.

Lemma 4.8. *Suppose R is a 3-ary relation in L and $f'(x) \le \frac{1}{2}(1/(x+1))$. Let $A \le C, T \in \mathcal{K}_f$ (with $A \ne C, T$), and let E be the free amalgam of C and T over A. Suppose $t_1, \dots, t_r \in T \setminus A$ are d-independent over A, and let $c \in C \setminus A$.*

Let $F = E \cup \{s_1, \ldots, s_r\}$ with additional relations $R(c, s_i, t_i)$ (for $1 \le i \le r$). Then $As_1 \cdots s_r, C, T \le F$ and $F \in \mathcal{K}_f$.

Proof. Suppose $C \subset V \subseteq F$. If $V \cap T = A$, then (by construction) $\delta(V) = \delta(C) + |V \setminus C|$; if $V \cap T \supset A$ then $\delta(V) \ge \delta(C) + \delta(V \cap T) - \delta(A) > \delta(C)$. In either case, $\delta(V) > \delta(C)$, so $C \le F$. A similar argument shows $T \le F$.

By free amalgamation, it is enough to prove the rest of the lemma in the case where $T = \mathrm{cl}_T(At_1 \cdots t_r)$ and $C = \mathrm{cl}_C(Ac)$. So henceforth assume this. Suppose $As_1 \cdots s_r \subset V \subseteq F$ has $\delta(V) \le \delta(As_1 \cdots s_r) = \delta(A) + r$. We can assume that $V \le F$. Clearly $c \in V$ and therefore $t_1, \ldots, t_r \in V$. It follows that $V = F$. But $\delta(F) = \delta(A) + r + 1$, a contradiction.

Finally we show that $F \in \mathcal{K}_f$. Let $X \subseteq F$. We need to show $\delta(X) \ge f(|X|)$. As $X \cap (T \cup C)$ is the free amalgamation of $X \cap T$ and $X \cap C$ over $X \cap A$, the structure X is of the same form as F (possibly together with some points s_i not lying in any relation in X). So it will suffice to prove that $\delta(F) \ge f(|F|)$.

Case 1: Suppose $|T \setminus A| \le r|C \setminus A|$.

Note that $|F| = |C| + |T \setminus A| + r$ and $\delta(F) = \delta(C) + r$. As $C \in \mathcal{K}_f$ we have $\delta(C) \ge f(|C|)$. Furthermore, as the graph of f lies below its tangent at any point, and $f'(x) \le \frac{1}{2}(1/(x+1)) \le 1/(x+1)$, we have

$$\begin{aligned} f(|F|) &\le f(|C|) + (|T \setminus A| + r) f'(|C|) \\ &\le f(|C|) + \frac{1}{(|C|+1)} r(|C \setminus A| + 1) \le \delta(C) + r = \delta(F), \end{aligned}$$

as required.

Case 2: Suppose $|T \setminus A| \ge r|C \setminus A|$.

This is similar. We have $|F| = |T| + |C \setminus A| + r$ and $\delta(F) = \delta(T) + 1$. Then

$$\begin{aligned} f(|F|) &\le f(|T|) + (|C \setminus A| + r) f'(|T|) \\ &\le f(|T|) + \frac{1}{2|T|}(|C \setminus A| + r) \le \delta(T) + 1 = \delta(F), \end{aligned}$$

using the fact that $|T \setminus A| \ge |C \setminus A|, r$. □

Proof of Proposition 4.5. Recall that we are assuming that the language L contains a 3-ary relation symbol R, so we can use the previous lemma. Let $A = \mathrm{acl}(\bar{a})$ and $B = \mathrm{acl}(Ab)$.

First, we note that it is enough to prove the proposition in the case where $\bar{y}$ is d-independent over $\bar{a}$ (that is, $d(\bar{y}/\bar{a}) = r$). To see this, take $\bar{y}_1 \subseteq \bar{y}$ which is d-independent over $\bar{a}$ and has $\bar{y} \in \mathrm{acl}(\bar{y}_1\bar{a})$; extend this to an r-tuple $\bar{y}'$ which is d-independent over $\bar{a}$. If $\bar{x} \in P^{r+2}$ has $\bar{y}_1 \in \mathrm{acl}(\bar{a}\bar{x})$, then $\bar{y} \in \mathrm{acl}(\bar{a}\bar{x})$.

Step 1: We first assume that $\bar{y} = (s_1, \ldots, s_r)$ is d-independent over $\bar{a}$ and $A\bar{y} \le M_f$. We shall show that there is $(b_0, \ldots, b_r) \in P^{r+1}$ with $\bar{y} \in \mathrm{acl}(\bar{a}, b_0, \ldots, b_r)$.

We apply Lemma 4.8 with T the free amalgam of r copies B_j $(1 \leq j \leq r)$ of B over A and C another copy of B. Let $b_1, \ldots, b_r, b_0$ be the corresponding copies of b (over A) inside $B_1, \ldots, B_r, C$ respectively. Let F be the disjoint union over A of $A\bar{y}$, C and T, but with the extra relations $R(b_j, s_j, b_0)$, where $1 \leq j \leq r$, as in the lemma. Then, by the lemma,

(i) $A\bar{y} \leq F$;

(ii) $B_j \leq F$; and

(iii) $F \in \mathcal{K}_f$.

Then by (i), (iii) and the $\leq$-extension property we can assume $F \leq M_f$; by (ii), we then have $\bar{x} = (b_0, b_1, \ldots, b_r) \in P^{r+1}$; then, because of the relations $R(b_j, s_j, b_0)$ we have $s_j \in \operatorname{acl}(b_0, b_j, A)$, so $\bar{y} \in \operatorname{acl}(\bar{a}\bar{x})$, as required.

Step 2: Now let $\bar{y} = (t_1, \ldots, t_r)$ be d-independent over A and let $T = \operatorname{acl}(A\bar{y})$. Let C be a copy of B over A with c the copy of b over A inside C, and let F be constructed as in the lemma. As in Step 1, we can assume that $F \leq M_f$. So $c \in P$ and $\bar{y} \in \operatorname{acl}(\bar{a}, c, s_1, \ldots, s_r)$. But by Step 1 (and $As_1 \cdots s_r \leq F$) the tuple $(s_1, \ldots, s_r)$ is in $\operatorname{acl}(\bar{a}\bar{z})$ for some $\bar{z} \in P^{r+1}$. The result follows. □

Acknowledgements

The author is very much indebted to Ehud Hrushovski for his comments and suggestions on a previous version of this paper. Thanks are also due to the referee for their helpful and detailed remarks.

References

[Chatzidakis et al. 1992] Z. Chatzidakis, L. van den Dries, and A. Macintyre, "Definable sets over finite fields", *J. Reine Angew. Math.* **427** (1992), 107–135. MR Zbl

[Elwes and Macpherson 2008] R. Elwes and D. Macpherson, "A survey of asymptotic classes and measurable structures", pp. 125–159 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

[Evans 2002] D. M. Evans, "$\aleph_0$-categorical structures with a predimension", *Ann. Pure Appl. Logic* **116**:1-3 (2002), 157–186. MR Zbl

[Hrushovski 1988] E. Hrushovski, "A stable $\aleph_0$-categorical pseudoplane", unpublished notes, 1988.

[Hrushovski 1997] E. Hrushovski, "Simplicity and the Lascar group", unpublished notes, 1997.

[Hrushovski 2012] E. Hrushovski, "Groupoids, imaginaries and internal covers", *Turkish J. Math.* **36**:2 (2012), 173–198. MR Zbl

[Hrushovski 2013] E. Hrushovski, "On pseudo-finite dimensions", *Notre Dame J. Form. Log.* **54**:3-4 (2013), 463–495. MR Zbl

[Hrushovski and Pillay 2011] E. Hrushovski and A. Pillay, "On NIP and invariant measures", *J. Eur. Math. Soc.* **13**:4 (2011), 1005–1061. MR Zbl

[Kallenberg 1997] O. Kallenberg, *Foundations of modern probability*, Springer, 1997. MR Zbl

[Kestner and Pillay 2011] C. Kestner and A. Pillay, "Remarks on unimodularity", *J. Symbolic Logic* **76**:4 (2011), 1453–1458. MR Zbl

[Macpherson and Steinhorn 2008] D. Macpherson and C. Steinhorn, "One-dimensional asymptotic classes of finite structures", *Trans. Amer. Math. Soc.* **360**:1 (2008), 411–448. MR Zbl

[Marimon 2022a] P. Marimon, "On the non-measurability of ω-categorical Hrushovski constructions", 2022. arXiv 2208.06323

[Marimon 2022b] P. Marimon, "Invariant Keisler measures for ω-categorical structures", 2022. arXiv 2211.14628

[Simon 2015] P. Simon, *A guide to NIP theories*, Lecture Notes in Logic **44**, Association for Symbolic Logic, Chicago, IL, 2015. MR Zbl

[Towsner 2010] H. Towsner, "A model theoretic proof of Szemerédi's theorem", 2010. arXiv 1002.4456

[Towsner 2018] H. Towsner, "An analytic approach to sparse hypergraphs: hypergraph removal", *Discrete Anal.* (2018), art. id. 3, 47 pages. MR Zbl

[Wagner 2000] F. O. Wagner, *Simple theories*, Mathematics and its Applications **503**, Kluwer Academic, Dordrecht, 2000. MR Zbl

Received 28 Feb 2022. Revised 12 Jul 2022.

DAVID M. EVANS:

david.evans@imperial.ac.uk
Department of Mathematics, Imperial College London, London, United Kingdom

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.255

Residue field domination in some henselian valued fields

Clifton Ealy, Deirdre Haskell and Pierre Simon

We generalize previous results about stable domination and residue field domination to henselian valued fields of equicharacteristic 0 with bounded Galois group, and we provide an alternate characterization of stable domination in algebraically closed valued fields for types over parameters in the field sort.

1. Introduction

The notion of domination of a type by its stable part was introduced and studied in the book [HHM 2008] and examined especially in the case of an algebraically closed valued field. The utility of the notion has been further demonstrated; for example, the space of stably dominated types in an algebraically closed valued field was analyzed in the book [Hrushovski and Loeser 2016] as an approach to understanding Berkovich spaces, and some structure theory has been developed for groups with a stably dominated generic type [Hrushovski and Rideau-Kikuchi 2019]. However, the stable part of a structure can seem like an unwieldy and abstract object. Since the stable sorts in an algebraically closed valued field are essentially those which are internal to the residue field, the intuition behind stable domination is that a stably dominated type is controlled by its trace in the residue field. By turning attention to the residue field instead of to the stable part, the hope is that this intuition could be used in two ways. The first is to develop a notion of domination that applies in more general valued fields in which the residue field is not necessarily stable. The second is to find a domination statement involving a simpler collection of sorts. This program was started in [Ealy et al. 2019], where we considered domination by sorts that are internal to the residue field in a real closed valued field. The present paper continues the project in the greater generality of henselian valued fields of equicharacteristic 0, provided that the Galois group is bounded. Details of the notation are given later; in the theorems quoted below, $\mathcal{U}$ is a monster model of the theory of valued fields in which we are working.

In our definition of residue field domination, we reduce the collection of sorts that are used for domination to the residue field itself, rather than the sorts that are

MSC2020: 03C60, 12J10, 12L12.
Keywords: valued fields, domination, model theory.

internal to the residue field. This may seem to be an unreasonably strong property, but we are able to show that it does hold in many cases, either assuming some algebraic conditions, or assuming stable domination, as in the following statements.

Theorem 4.5. *Let $C \subseteq \mathcal{U}$ be a subfield and let a be a (possibly infinite) tuple of field elements such that the field generated by Ca is an unramified extension of C with the good separated basis property over C, and such that $k(Ca)$ is a regular extension of $k(C)$. Then* $\mathrm{tp}(a/C)$ *is residue field dominated.*

Theorem 4.6. *Let $C \subseteq \mathcal{U}$ be a subfield, let $a \in \mathcal{U}$, and let $\widetilde{\mathcal{U}}$ be the algebraic closure of $\mathcal{U}$. Assume that* $\mathrm{tp}(a/C)$ *is stably dominated in the structure $\widetilde{\mathcal{U}}$. Then in the structure $\mathcal{U}$,* $\mathrm{tp}(a/C^+)$ *is residue field dominated, where* $C^+ = \mathrm{acl}(C) \cap \mathrm{dcl}(Ca)$.

There are, however, important examples, when the base of a type is not in the field sort, where stable domination does not reduce to residue field domination. For instance, a major theme of stable domination is that types (with a few caveats) are always stably dominated over the value group. However, they need not be residue field dominated over the value group. In addition to the residue field, one needs information from sorts that are internal to the residue field. These turn out to be given by fibers of the valuation map in RV. We thus introduce another notion, RV-domination, and show that types are RV-dominated over their value groups.

Theorem 3.11. *Let L, M be subfields of $\mathcal{U}$ with $C \subseteq L \cap M$ a valued subfield. Assume that $k(L)$ is a regular extension of $k(C)$, $\Gamma_L \subseteq \Gamma_M$, Γ_L / Γ_C is torsion free and that L has the good separated basis property over C. Then* $\mathrm{tp}(L/C\Gamma_L)$ *is* RV*-dominated.*

An important insight of this paper is that one key step in proving domination results is the existence of a separated basis. This insight allows us to distinguish between purely algebraic concepts and the more model-theoretic ones. In particular, we derive the following algebraic characterization of stable domination for types in the field sort in an algebraically closed valued field.

Theorem 3.6. *Suppose that $\mathcal{U}$ is algebraically closed. Let $C \subset \mathcal{U}$ be a subfield, let a be a tuple of valued field elements, and let L be the definable closure of Ca in the valued field sort. Assume L is a regular extension of C. Then the following are equivalent.*

(i) $\mathrm{tp}(a/C)$ *is stably dominated.*

(ii) *L has the good separated basis property over C and L is an unramified extension of C.*

When restricted to the main sort, the domination statements can be given a purely valuation-theoretic form, as asserting the existence of automorphisms under certain hypotheses; these are Proposition 3.1 and Theorem 3.10.

In the time since this paper was originally submitted, further work has been done by several authors. We mention in particular the work of Vicaria [2021], which uses, and to some extent generalizes, the results of this paper. She does not need the hypothesis that the Galois group of the field is bounded. However, she uses a rather different language, with sorts for the cosets of the subgroups of the n-th powers in RV. Also relevant is the work of Cubides Kovacsics, Hils and Ye [Cubides Kovacsics et al. 2021], which independently obtains type implication results using the existence of a separated basis (there called being vs-defectless).

The outline of the paper is as follows. In the remainder of the introduction we state a quantifier elimination result for the theory in which we work, give the definition of domination and some associated properties, and recall some elementary properties of type implication and regular field extensions. In Section 2, we define the notion of a good separated basis over a base field C and some consequences, in particular the relation to the assumption that C is a maximal field. In Section 3, we prove some preliminary results towards residue field domination, using the separated basis hypothesis. Finally, in Section 4 we derive the full domination results, after showing that the geometric sorts can be resolved in the field sort.

Notation. We work in two languages, $\mathcal{L}$ and $\widetilde{\mathcal{L}}$, and two structures, $\mathcal{U}$ and $\widetilde{\mathcal{U}}$.

We fix K, a henselian valued field of equicharacteristic 0 with bounded Galois group. The first language, $\mathcal{L}$, is described in Proposition 1.3 below; it depends on K. We fix the theory T of K in the language $\mathcal{L}$. We let $\mathcal{U}$ be a monster model of T.

The second language, $\widetilde{\mathcal{L}}$, is the language often used for algebraically closed valued fields. We equip the field sort with the usual ring language and use the notation k for the residue field sort in the usual ring language, Γ for the value group sort in the language of ordered abelian groups and RV for the RV sort with the induced multiplicative group structure. We include the geometric sorts required to eliminate imaginaries, namely $\bigcup_{n=1}^{\infty} S_n$ for the lattices and $\bigcup_{n=1}^{\infty} T_n$ for their torsors. However, the resolution results of Theorem 4.2 below and Chapter 11 of [HHM 2008] allow us to avoid working with the geometric sorts directly in this paper, and thus we omit their (rather lengthy) definition; a detailed description can be found in [HHM 2006, Section 3.1; 2008, Section 7.4].

We let $\widetilde{\mathcal{U}}$ be a monster model of ACVF such that the field sort of $\mathcal{U}$ embeds into the field sort of $\widetilde{\mathcal{U}}$, and such that every automorphism of $\mathcal{U}$ extends to an automorphism of $\widetilde{\mathcal{U}}$ (e.g., $\widetilde{\mathcal{U}}$ could be the algebraic closure of $\mathcal{U}$). Throughout the paper, we use a subscript $\widetilde{\mathcal{L}}$ to indicate not just that we are working in the language $\widetilde{\mathcal{L}}$, but that we are also working in the algebraically closed valued field $\widetilde{\mathcal{U}}$ (for instance, when taking definable closure, or specifying a type); no subscript indicates that we are working in the language $\mathcal{L}$ and in $\mathcal{U}$.

Given any definable set $\mathcal{S}$ and set of parameters C, we write $\mathcal{S}(C) = \mathrm{dcl}(C) \cap \mathcal{S}$. If C is a substructure of $\mathcal{U}$, we write $\mathcal{S}_C = C \cap \mathcal{S}$. For any field, we use the superscript alg to denote its field-theoretic algebraic closure. On any field, and in particular on the residue field k, we have an independence relation $\mathop{\smile\!\!\!\!\!\!\mid}\nolimits^{\mathrm{alg}}$: for $A, B \subseteq k$, $A \mathop{\smile\!\!\!\!\!\!\mid}\nolimits^{\mathrm{alg}}_C B$ means that any finite subset of $k(AC)$ that is field algebraically independent over $k(C)$ remains so over $k(BC)$.

Quantifier elimination. The language $\mathcal{L}$ is chosen so that the theory of the valued field that we are working with has quantifier elimination. This is derived from the following results as described below. The first is a result of Chernikov and Simon translated into the notation of valued fields. Note that bounded Galois group implies that the n-th powers have finite index in the field [Fehm and Jahnke 2016] and hence also in RV. This is our paper's only use of the assumption of bounded Galois group. One may construct henselian fields of equicharacteristic 0 where n-th powers have finite index in RV but which do not have bounded Galois group [Fehm and Jahnke 2016, Proposition 5.1]. Our results apply to these fields as well.

Fact 1.1 [Chernikov and Simon 2019, Proposition 3.1]. *Let K be a henselian valued field of equicharacteristic* 0 *with bounded Galois group. Assume the language $\mathcal{L}$ is chosen so that*

- RV *has its multiplicative group structure*, *a predicate for k as a multiplicative subgroup*, *n-th power predicates*, *constants naming a countable subgroup containing representatives of the* (*finitely many*) *cosets of the n-th powers for $n < \omega$* (*where representatives of classes which intersect k are chosen in k*), *a sort for Γ*, *and a map $v : \mathrm{RV} \to \Gamma$*;
- *the language of Γ expands the structure induced from K*, *has no function symbols apart from $+$*, *and eliminates quantifiers*;
- *the language of k expands the structure induced from K*, *has no function symbols apart from $\cdot$*, *and eliminates quantifiers.*

Then (RV, Γ, k) *has quantifier elimination.*

Fact 1.2 [Pas 1989, Theorem 4.1]. *Let T be the theory of a henselian valued field of equicharacteristic* 0, *in the language with sorts for k and Γ*, *expanded by the angular component map. Then T has elimination of field quantifiers.*

One can show (e.g., [Cluckers and Loeser 2007; Rideau-Kikuchi 2017, Theorem A; Scanlon 2003, Corollary 5.8, assuming the trivial derivation]) that elimination of field quantifiers with an angular component map implies elimination of field quantifiers relative to RV. In our case, RV itself eliminates quantifiers as in Fact 1.1, and thus we may conclude Proposition 1.3 below. We remark that the form in which this proposition is generally used is the following: if $A, B \subset \mathcal{U}$ are valued fields, and $\sigma : A \to B$ is a valued field isomorphism which induces an isomorphism of RV-structures $\mathrm{RV}_A \to \mathrm{RV}_B$, then σ extends to an automorphism of $\mathcal{U}$.

Proposition 1.3. *Let K be a henselian valued field of equicharacteristic* 0 *with bounded Galois group. Work in the language with*

- *the language of rings on K,*
- *a sort for* RV *and a sort for* Γ, *each in the language of groups,*
- *a predicate for $k \subset$* RV,
- *a map* $\mathrm{rv} : K \to \mathrm{RV}$,
- *a map* $v : \mathrm{RV} \to \Gamma$,
- *predicates for every subset of k^m and Γ^m definable without parameters in the structure induced from K,*
- *predicates for the n-th powers in* RV, *and*
- *constants for a countable subgroup of* RV *containing coset representatives for each of the n-th power subgroups of* RV, *chosen in k where possible.*

Then K has quantifier elimination.

Remark 1.4. It follows from this proposition that the value group and residue field are stably embedded in the following strong form: if $\varphi(x, a)$ defines a subset of k^n, then there is a term t and quantifier-free formula θ such that $\theta(x, t(a))$ defines the same subset. Given that θ is quantifier free, it is clear that $t(a)$ lies in the RV-structure (either in RV itself or in Γ). It is easy to check that $t(a)$ can be chosen to lie in the residue field. The same argument also shows that if X is a subset of Γ defined over a then it is also defined over $t(a) \in \Gamma$ for some term t. Note that this is slightly stronger than the definition of stable embeddedness, which does not require the parameter in the stably embedded set to be in $\mathrm{dcl}(a)$.

We would not in general expect this strong form of stable embeddedness to hold for an individual fiber in RV, which we write as $\mathrm{RV}_\gamma = \{x \in \mathrm{RV} : v(x) = \gamma\}$. For consider the subset of $\mathrm{RV}_\gamma \times \mathrm{RV}_\gamma$ defined by $x \cdot y^{-1} = a$, where $a \in k$. However, if one assumes that RV_γ contains some point a_0 that is expressible as a term $t_0(a)$, then it is again true that any definable subset of ${\mathrm{RV}_\gamma}^n$ defined over a is defined over a term $t(a)$ with $t(a) \in \mathrm{RV}_\gamma$. For if X is such a set, $X \cdot {a_0}^{-1}$ is a definable subset of the residue field, and therefore definable over $t'(a) \in k$ for some term t'. Hence $X \cdot {a_0}^{-1}$ is also definable over $t'(a) \cdot t_0(a) \in \mathrm{RV}_\gamma$, and so is X.

Lastly, the quantifier elimination result implies that the residue field and value group are orthogonal to each other.

Domination: definition and basic properties. Residue field domination is defined by analogy with stable domination, which we now recall [HHM 2008, Definition 3.9].

Given a set of parameters C in $\widetilde{\mathcal{U}}$, let St_C be the multisorted structure whose sorts are the C-definable stable, stably embedded subsets of $\widetilde{\mathcal{U}}$. The structure St_C is itself stable, so stable forking gives an independence relation $\mathop{\smile\!\!\!\!\!\!\mid}$.

Definition 1.5. We say that $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/C)$ is *stably dominated* if for any $b \in \widetilde{\mathcal{U}}$, whenever $\mathrm{St}_C(aC) \mathop{\smile\!\!\!\!\!\!\mid}_C \mathrm{St}_C(bC)$ we have $\mathrm{tp}_{\widetilde{\mathcal{L}}}(b/C\mathrm{St}_C(aC)) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(b/Ca)$.

The definition captures our intuition that a stably dominated type should have no interaction with the value group in the following sense.

Fact 1.6 [HHM 2008, Corollary 10.8]. *The type* $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/C)$ *is stably dominated if and only if it is orthogonal to* Γ.

Notice that Corollary 10.8 and the definition of orthogonality in [HHM 2008, Definition 10.1] are only given in the original for the case when a is a unary sequence. However they both can be stated in more generality, since for any element s and any set C in the geometric sorts of a valued field, there is a unary sequence, a, with the same $\widetilde{\mathcal{L}}$-definable closure over C [HHM 2006, Proposition 2.3.10; 2008, Proposition 7.14]. For such an s and a, one may define $\mathrm{tp}(s/C)$ to be orthogonal to Γ if $\mathrm{tp}(a/C)$ is orthogonal to Γ, noting by [HHM 2008, Lemma 10.9] that this is independent of the choice of a.

The structure St_C can be defined in any structure, but it may be trivial or hard to identify. In an algebraically closed valued field, St_C is interdefinable with the collection of sorts internal to the residue field, which are themselves interdefinable (with parameters) with the residue field. This motivates the following definition for a valued field that is not necessarily algebraically closed. Notice that residue field domination as defined here is a very strong property, since the independence notion we are working with is very weak. It is thus surprising that we can prove instances of residue field domination in Section 4.

Definition 1.7. We say that $\mathrm{tp}(a/C)$ is *residue field dominated* if for any $b \in \mathcal{U}$, if $k(aC) \mathop{\smile\!\!\!\!\!\!\mid}_C^{\mathrm{alg}} k(bC)$, then $\mathrm{tp}(b/Ck(Ca)) \vdash \mathrm{tp}(b/Ca)$.

When $\mathcal{U}$ is itself algebraically closed, it is immediate that residue field domination implies stable domination. If $\mathcal{U}$ is, for example, a real closed valued field, this implication does not hold. The converse is not true even when $\mathcal{U}$ is algebraically closed, as the following example illustrates. In particular, this example shows that issues may arise when the type is over parameters in the value group sort.

Example 1.8. Let $C = \mathbb{Q}$ and let $a \in \mathcal{U}$ be a field element of positive valuation. Then C is maximal because it is trivially valued, $L = \mathrm{dcl}(a)$ has $k_L = k_C$ and hence is automatically a regular extension, and Γ_L is a torsion-free extension of Γ_C (which is the trivial group). So by [HHM 2006, Theorem 12.18], $\mathrm{tp}(a/C\Gamma_L)$ is stably dominated. However, $\mathrm{tp}(a/C\Gamma_L)$ is not residue field dominated. For if we take $M = L$, the independence condition holds trivially since $k_M = k_L = k_C$, but it is not the case that $\mathrm{tp}(L/C\Gamma_L k_L)$ implies $\mathrm{tp}(L/M) = \mathrm{tp}(L/L)$.

We are able to prove a version of [HHM 2006, Theorem 12.18], involving RV-domination instead of residue field domination, which we define in Definition 3.8.

In [HHM 2006], it is shown that stable domination is insensitive to whether or not the base is algebraically closed.

Fact 1.9 [HHM 2006, Corollary 3.31]. *The type* $\mathrm{tp}(a/C)$ *is stably dominated if and only if* $\mathrm{tp}(a/\operatorname{acl}(C))$ *is stably dominated.*

This is not true for residue field domination, as the following example illustrates. We make use here, and many times later, of the following basic fact.

Fact 1.10. *Let* $C \subset \widetilde{\mathcal{U}}$, $a \in \widetilde{\mathcal{U}}$. *Then* $\operatorname{dcl}_{\widetilde{\mathcal{L}}}(Ca)$ (*restricted to the field sort*) *is the henselization of the field generated by a over* C.

Example 1.11. Let K be an algebraically closed valued field of characteristic 0, let t be an element of positive valuation, and consider $C = \operatorname{dcl}(\mathbb{Q}(t))$. We note that $\sqrt{t}$ cannot be in C since the definable closure of $\mathbb{Q}(t)$ is the henselization of $\mathbb{Q}(t)$, which is an immediate extension. Let $a = \sqrt{t}$. Clearly $\mathrm{tp}(a/\operatorname{acl}(C))$ is stably dominated and residue field dominated. Yet $\mathrm{tp}(a/C)$ is stably dominated but not residue field dominated. To see the second statement, choose $b = a$. One has $k(aC) \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.9em\lower0.2ex\hbox{$|$}}_C^{\mathrm{alg}} k(aC)$ since $a \in \operatorname{acl}(C)$. Since $\sqrt{t}$ generates a ramified extension of C, $k(Ca) = k(C)$. Thus $\mathrm{tp}(a/Ck(Ca)) = \mathrm{tp}(a/C)$, and clearly $\mathrm{tp}(a/C)$ cannot imply $\mathrm{tp}(a/Ca)$.

On the other hand, $\mathrm{tp}(a/C)$ is stably dominated. Since $a \in \operatorname{acl}(C)$, a is in a C-definable stable, stably embedded set, i.e., is in $\mathrm{St}(C)$. So automatically $\mathrm{tp}(b/C\,\mathrm{St}_C(a))$ implies $\mathrm{tp}(b/Ca)$ for any b.

However we do get the following, slightly weaker, statement. The proof uses Proposition 1.15 below.

Proposition 1.12. *For* $C \subset \mathcal{U}$ *and* $a \in \mathcal{U}$, *let* $C^+ = \operatorname{acl}(C) \cap \operatorname{dcl}(Ca)$. *Then* $\mathrm{tp}(a/C^+)$ *is residue field dominated if and only if* $\mathrm{tp}(a/\operatorname{acl}(C))$ *is residue field dominated.*

Proof. For the right-to-left direction, choose b such that $k(C^+a) \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.9em\lower0.2ex\hbox{$|$}}_{C^+}^{\mathrm{alg}} k(C^+b)$. Since fields code finite sets, if $d_1 \in \operatorname{acl}(C)$ and the orbit of d_1 over C is $d_1, \dots, d_n$, then $\{d_1, \dots, d_n\} \in \operatorname{dcl}(C)$ and $d_1, \dots, d_n \in \operatorname{alg}(\operatorname{dcl}(C))$, where alg denotes the field-theoretic algebraic closure. Thus $\operatorname{acl}(C) \subseteq \operatorname{alg}(C^+)$. Note that we have the implications

$$\begin{aligned} k(C^+a) \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.9em\lower0.2ex\hbox{$|$}}_{C^+}^{\mathrm{alg}} k(C^+b) &\Longrightarrow \operatorname{alg}(k(C^+a)) \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.9em\lower0.2ex\hbox{$|$}}_{\operatorname{alg}(C^+)}^{\mathrm{alg}} \operatorname{alg}(k(C^+b)) \\ &\Longrightarrow k(\operatorname{alg}(C^+a)) \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.9em\lower0.2ex\hbox{$|$}}_{\operatorname{alg}(C^+)}^{\mathrm{alg}} k(\operatorname{alg}(C^+b)) \\ &\Longrightarrow k(\operatorname{acl}(C)a) \mathrel{\raise0.2ex\hbox{$\smile$}\kern-0.9em\lower0.2ex\hbox{$|$}}_{\operatorname{acl}(C)}^{\mathrm{alg}} k(\operatorname{acl}(C)b). \end{aligned}$$

Therefore, we have $\operatorname{tp}(b/\operatorname{acl}(C)k(\operatorname{acl}(C)a)) \vdash \operatorname{tp}(b/\operatorname{acl}(C)a)$, and we want $\operatorname{tp}(b/C) \vdash \operatorname{tp}(b/C^+a)$. Choose $\varphi(x,a) \in \operatorname{tp}(b/C^+a)$. This is implied by some $\psi(x,c,d) \in \operatorname{tp}(b/\operatorname{acl}(C)k(\operatorname{acl}(C)a))$, with $c \in \operatorname{acl}(C)$ and $d \in k(\operatorname{acl}(C)a)$. Let $X = \{\sigma(c)\sigma(d) : \sigma \in \operatorname{Aut}(\mathcal{U}/C^+a)\}$. Notice that $X_1 = \{\sigma(c) : \sigma \in \operatorname{Aut}(\mathcal{U}/C^+a)\}$ is finite, so Ca-definable, and in $\operatorname{acl}(C)$, hence fixed by any automorphism fixing C^+. Also $X_2 = \{\sigma(d) : \sigma \in \operatorname{Aut}(\mathcal{U}/C^+a)\}$ is C^+a-definable and in the residue field, and thus $X_2 \in k(C^+a)$.

Thus, the formula θ_0 given by

$$\bigvee_{\sigma(c)\in X_1} \bigvee_{\sigma'(d)\in X_2} \psi(x,\sigma(c),\sigma'(d))$$

is over $C^+k(C^+a)$ as desired, and for any $\sigma(c)\sigma(d)$ in X, we have $\psi(x,\sigma(c),\sigma(d))$ implies $\varphi(x,a)$. However, if σ' is some other automorphism fixing C^+a, it may be the case that $\psi(x,\sigma(c),\sigma'(d))$ does not imply $\varphi(x,a)$, and so we must tweak θ_0. If σ' is such an isomorphism, then $\sigma(c)\sigma'(d) \not\equiv_{C^+a} \sigma(c)\sigma(d)$ and thus $\sigma'(d) \not\equiv_{\sigma(c)k(C^+a)} \sigma(d)$. For each $\sigma \in \operatorname{Aut}(\mathcal{U}/C^+a)$, let $e_{\sigma(c)}$ be the orbit of $\sigma(d)$ over $\sigma(c)k(C^+a)$. Then the formula, θ, given by

$$\bigvee_{\sigma(c)\in X_1} \bigvee_{d'\in e_{\sigma(c)}} \psi(x,\sigma(c),d')$$

implies $\varphi(x,a)$.

We claim that $\{\sigma(c)e_\sigma(c) : \sigma \in \operatorname{Aut}(\mathcal{U}/C^+a)\}$ is $C^+k(C^+a)$-definable, and hence the displayed formula above gives the required domination statement. Consider τ an automorphism fixing $C^+k(C^+a)$. Since τ fixes C^+, τ maps X_1 to itself, so there is an automorphism σ fixing C^+a such that $\tau(c) = \sigma(c)$. It suffices to show that $\tau(d) \in e_{\sigma(c)}$. By definition, $\sigma(d) \in e_{\sigma(c)}$. Now $\tau \circ \sigma^{-1}$ fixes $\sigma(c)$ and $k(C^+a)$, and $\tau \circ \sigma^{-1}(\sigma(d)) = \tau(d)$, which hence lies in the $\operatorname{Aut}(\mathcal{U}/\sigma(c)k(C^+a))$-orbit of $\sigma(d)$, as required.

For the other direction, take b with $k(\operatorname{acl}(C)a) \mathop{\smile\!\!\!\!\!|}^{\mathrm{alg}}_{\operatorname{acl}(C)} k(\operatorname{acl}(C)b)$. It suffices, by Proposition 1.15, to show that $\operatorname{tp}(a/\operatorname{acl}(C)k(\operatorname{acl}(C)b)) \vdash \operatorname{tp}(a/\operatorname{acl}(C)b)$. Note that, by replacing the set $k(\operatorname{acl}(C)a)$ with a subset and replacing the set $\operatorname{acl}(C)$ in the base with something interalgebraic with it, we have

$$k(C^+a) \mathop{\smile\!\!\!\!\!|}^{\mathrm{alg}}_{C^+} k(\operatorname{acl}(C)b).$$

Thus we may apply residue field domination of $\operatorname{tp}(a/C^+)$, where our tuple from $\mathcal{U}$ is $\operatorname{acl}(C)b$, obtaining (again applying Proposition 1.15)

$$\operatorname{tp}(a/C^+k(\operatorname{acl}(C)b)) \vdash \operatorname{tp}(a/\operatorname{acl}(C)b).$$

So certainly

$$\operatorname{tp}(a/\operatorname{acl}(C)k(\operatorname{acl}(C)b)) \vdash \operatorname{tp}(a/\operatorname{acl}(C)b)$$

as well. □

Type implications. Since many of our arguments involve showing type implications, it is useful to make the following very general observations.

Lemma 1.13. *Let A, B, C be subsets of a monster model $\mathcal{U}$ in some language, with $C \subseteq A \cap B$. Then*

(i) $\operatorname{tp}(A/C) \vdash \operatorname{tp}(A/B)$ *is equivalent to* $\operatorname{tp}(B/C) \vdash \operatorname{tp}(B/A)$;

(ii) *if* $\operatorname{tp}(A/C) \vdash \operatorname{tp}(A/B)$ *and* $\operatorname{tp}(B'/C) = \operatorname{tp}(B/C)$, *then* $\operatorname{tp}(A/C) \vdash \operatorname{tp}(A/B')$.

Proof. (i) Suppose that $\operatorname{tp}(A/C) \vdash \operatorname{tp}(A/B)$ and $\operatorname{tp}(B'/C) = \operatorname{tp}(B/C)$. Let $\sigma \in \operatorname{Aut}(\mathcal{U}/C)$ with $\sigma(B') = B$. As $\operatorname{tp}(\sigma(A)/C) = \operatorname{tp}(A/C)$, by the type implication assumption, also $\operatorname{tp}(\sigma(A)/B) = \operatorname{tp}(A/B)$. Thus there is $\tau \in \operatorname{Aut}(\mathcal{U}/B)$ such that $\tau(\sigma(A)) = A$. Then $\tau(\sigma(B')) = B$, so $\operatorname{tp}(B'/A) = \operatorname{tp}(B/A)$.

(ii) By (i), it is equivalent to show that $\operatorname{tp}(B'/C) \vdash \operatorname{tp}(B'/A)$, which is the same statement as $\operatorname{tp}(B/C) \vdash \operatorname{tp}(B'/A)$. Also by (i), we have $\operatorname{tp}(B/C) \vdash \operatorname{tp}(B/A)$. So we need only establish that $\operatorname{tp}(B'/A) = \operatorname{tp}(B/A)$. But since we know that $\operatorname{tp}(B/C) \vdash \operatorname{tp}(B/A)$, we know that anything (e.g., B') that realizes $\operatorname{tp}(B/C)$ must also realize $\operatorname{tp}(B/A)$. Thus $B' \models \operatorname{tp}(B/A)$ and $\operatorname{tp}(B'/A) = \operatorname{tp}(B/A)$. □

The following lemma is stated in [HHM 2008, Remark 3.7] for the stable part of a structure. We prove it here using Remark 1.4 which allows us to avoid the assumption of elimination of imaginaries. Let $\mathcal{S}$ be any definable set that is stably embedded in the strong sense defined in Remark 1.4. Later we will take $\mathcal{S}$ to be the residue field, the value group, or some collection of fibers of RV, where for each γ, $\mathrm{RV}_\gamma(CB)$ is nonempty.

Lemma 1.14. *For any sets A, B, C in $\mathcal{U}$*, $\operatorname{tp}(B/C\mathcal{S}(CB)) \vdash \operatorname{tp}(B/C\mathcal{S}(CB)\mathcal{S}(CA))$.

Proof. We may assume B is finite. Take $B' \equiv_{C\mathcal{S}(CB)} B$. We wish to show that $B' \equiv_{C\mathcal{S}(CB)\mathcal{S}(CA)} B$, so take $\varphi(x, a, b) \in \operatorname{tp}(B/C\mathcal{S}(CA)\mathcal{S}(CB))$ with $a \in \mathcal{S}(CA)$ and $b \in \mathcal{S}(CB)$. We wish to show that $\varphi(B', a, b)$ holds.

Consider the set defined by $\varphi(B, y, b)$. This is a subset of $\mathcal{S}$, defined over CB, and hence definable by some $\theta(y, \tilde{b})$, where $\tilde{b} \in \mathcal{S}(CB)$ as described in Remark 1.4. Thus $\forall y\, [\theta(y, \tilde{b}) \rightarrow \varphi(x, y, b)] \in \operatorname{tp}(B/C\mathcal{S}(CB))$.

Since $\forall y\, [\theta(y, \tilde{b}) \rightarrow \varphi(B', y, b)]$ holds and $\theta(a, \tilde{b})$ also holds, it follows that $\varphi(B', a, b)$ holds. □

From this, we derive equivalences for the type implication in the definition of residue field domination.

Proposition 1.15. *For any a, b, C in $\mathcal{U}$ the following are equivalent*:

(i) $\operatorname{tp}(b/C\mathcal{S}(Ca)) \vdash \operatorname{tp}(b/Ca)$.

(ii) $\operatorname{tp}(a/C\mathcal{S}(Cb)) \vdash \operatorname{tp}(a/Cb)$.

(iii) $\operatorname{tp}(\mathcal{S}(bC)/C\mathcal{S}(Ca)) \cup \operatorname{tp}(b/C) \vdash \operatorname{tp}(b/Ca)$.

(iv) $\mathrm{tp}(a/C\mathcal{S}(Ca)\mathcal{S}(Cb)) \vdash \mathrm{tp}(a/Cb)$.

(v) $\mathrm{tp}(a/C\mathcal{S}(Ca)) \vdash \mathrm{tp}(a/Cb)$.

Proof. The proof of the equivalence of (i), (ii) and (iii) is exactly the proof of [HHM 2008, Lemma 3.8], replacing the stable, stably embedded sorts with the definable set $\mathcal{S}$, and referring to Lemma 1.14 in lieu of [HHM 2008, Remark 3.7]. The fact that (ii) implies (iv) is trivial, and that (iv) implies (v) is immediate by Lemma 1.14.

To prove that (v) implies (i), assume (v). Take $b, b' \models \mathrm{tp}(b/C\mathcal{S}(Ca))$ and σ witnessing this. Suppose that $\sigma^{-1}(a) = \tilde{a}$ and note that $a, \tilde{a} \models \mathrm{tp}(a/C\mathcal{S}(Ca))$, and thus by (v) they both satisfy $\mathrm{tp}(a/Cb)$. Choose $\tau : a \mapsto \tilde{a}$ witnessing this. Thus $(\sigma \circ \tau)(a) = \sigma(\tilde{a}) = a$ and $(\sigma \circ \tau)(b) = \sigma(b) = b'$, and we have $b, b' \models \mathrm{tp}(b/Ca)$. □

We will have need of the following result, which we will use in the form of the subsequent lemma.

Fact 1.16 [HHM 2008, Proposition 8.22(ii)]. *Let $C \subseteq A$, B be algebraically closed valued fields and suppose that $\Gamma(C) = \Gamma(A)$, the transcendence degree of B over C is* 1, *and there is no embedding of B into A over C. Then $\Gamma(AB) = \Gamma(B)$.*

Recall that we use $\Gamma(C)$ to mean $\mathrm{dcl}(C) \cap \Gamma$. In the following lemma, as we are working in $\widetilde{\mathcal{U}}$, the definable closure is taken in $\widetilde{\mathcal{L}}$.

Lemma 1.17. *Let $C \subseteq F$, L be valued fields contained in $\widetilde{\mathcal{U}}$ such that L is transcendence degree at least* 1 *over C, $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$, and $\Gamma(F) = \Gamma(C)$. Then $\Gamma(LF) = \Gamma(L)$.*

Proof. We proceed by induction on the transcendence degree n of L over C.

Assume $n = 1$. Since $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$, no $\ell \in L \setminus \mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$ can be embedded into $\mathrm{acl}_{\widetilde{\mathcal{L}}}(F)$ over C. For suppose that $\ell \equiv_C \ell'$. Then also $\ell \equiv_F \ell'$. If ℓ' could be chosen in $\mathrm{acl}_{\widetilde{\mathcal{L}}}(F)$, then ℓ would be an element of the finite set of elements realizing $\mathrm{tp}_{\widetilde{\mathcal{L}}}(\ell'/F)$. But this applies equally to any element of $\mathrm{tp}_{\widetilde{\mathcal{L}}}(\ell/C)$, and hence this type has finitely many realizations. Then ℓ would be in $\mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$. Hence there is no embedding of $\mathrm{acl}_{\widetilde{\mathcal{L}}}(L)$ into $\mathrm{acl}_{\widetilde{\mathcal{L}}}(F)$ over $\mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$, and we apply Fact 1.16 to obtain $\Gamma(\mathrm{acl}_{\widetilde{\mathcal{L}}}(L)\,\mathrm{acl}_{\widetilde{\mathcal{L}}}(F)) = \Gamma(\mathrm{acl}_{\widetilde{\mathcal{L}}}(L))$. Recalling that we have defined $\Gamma(A)$ to be the definable closure of the value group of A, we have $\Gamma(LF) = \Gamma(L)$.

Assume the result for $m < n$ and suppose L has transcendence degree n over C. Let $C \subseteq C' \subseteq L$ be such that L is transcendence degree 1 over C'. Note that $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C') \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/FC')$, since

$$\ell \equiv_{C'} \ell' \Longrightarrow \ell C' \equiv_C \ell' C' \Longrightarrow \ell C' \equiv_F \ell' C' \Longrightarrow \ell \equiv_{C'F} \ell'.$$

Thus, by our inductive hypothesis, $\Gamma(C'F) = \Gamma(C')$. Now one may repeat the argument of the $n = 1$ case with C' playing the role of C and $C'F$ playing the role of F. □

Regular extensions. The following three properties of regular extensions of fields are implicit in many of our arguments.

Fact 1.18 [Lang 2002, VIII, 4.12]. *Suppose C is a field, L is a regular field extension of C and M is any field extension of C, all contained in $\widetilde{\mathcal{U}}$. Then $L \mathop{\smile\!\!\!\!\!\!\perp}^{\mathrm{alg}}_{C} M$ implies L and M are linearly disjoint over C.*

Lemma 1.19. *Let C and L be valued fields contained in $\widetilde{\mathcal{U}}$ such that $C \subseteq L$ is a regular extension of fields and L is henselian. Then $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/\mathrm{acl}_{\widetilde{\mathcal{L}}}(C))$.*

Proof. Note that we may restrict our attention to the valued field sort of $\widetilde{\mathcal{U}}$. Let $a \in L$ be a finite tuple. Let X be an $\mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$-definable set containing a and let $X = X_1, \dots, X_n$ be the conjugates of X over C. We may assume that the X_i are pairwise disjoint (consider the boolean algebra generated by the X_i and replace X by the atom containing a).

Suppose that X_1 is defined by $\varphi(x, b)$ with $b \in \mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$. Consider the set B of conjugates $\{b = b_1, \dots, b_k\}$ of b over C, noting that k could be larger than n. Let S_1 be the subset of B consisting of those b_i such that $\varphi(x, b_i)$ defines X_1. Since fields code finite sets, there is a tuple $d_1 \in \mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$ that is a code for S_1. Consider the conjugates $D = \{d_1, \dots, d_n\}$ of d_1 over C. Note that X_1 is definable over d_1, so it suffices to show that $d_1 \in C$.

Since D is $\widetilde{\mathcal{L}}$-definable over C, d_1 is $\widetilde{\mathcal{L}}$-definable over Ca. Since in an algebraically closed valued field of characteristic 0, the definable closure of a set of field elements is the henselization of the field generated by those elements, d_1 is in the henselian closure of Ca, which is included in L. Since L is a regular extension of C and d_1 is algebraic over C, we conclude $d_1 \in C$ and hence X is $\widetilde{\mathcal{L}}$-definable over C. □

Lemma 1.20. *Let C, F and L be valued fields contained in $\widetilde{\mathcal{U}}$ such that $C \subseteq F \cap L$, L is a regular extension of F, $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$, and C is not trivially valued. Then L and F are linearly disjoint over C.*

Proof. By Lemma 1.13(i), since $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$, also $\mathrm{tp}_{\widetilde{\mathcal{L}}}(F/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(F/L)$. Suppose that there are $\vec{\ell} \in L$ and $\vec{f} \in F$ such that $\vec{\ell} \cdot \vec{f} = 0$ with $\vec{\ell}, \vec{f} \neq 0$, and let $\varphi(\vec{x}, \vec{\ell})$ express this of $\vec{x}$. As $\varphi(\vec{x}, \vec{\ell}) \in \mathrm{tp}_{\widetilde{\mathcal{L}}}(F/L)$, it is implied by some formula $\psi(\vec{x}, c) \in \mathrm{tp}_{\widetilde{\mathcal{L}}}(F/C)$. As $\mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$ is a model, there is some $\vec{d} \in \mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$ such that $\psi(\vec{d}, c)$. Hence, $\varphi(\vec{d}, \vec{\ell})$ holds, i.e., $\vec{\ell} \cdot \vec{d} = 0$ and $\vec{d} \neq 0$. Note that $C \subseteq L$ is a regular extension of fields (in characteristic 0) if and only if L is linearly disjoint from $\mathrm{acl}_{\widetilde{\mathcal{L}}}(C)$ over C. So there must also be $\vec{c} \in C$ with $\vec{\ell} \cdot \vec{c} = 0$. □

2. Separated bases

The notion of a good separated basis was isolated in [HHM 2008], based on earlier observations by many different authors. In this section, we show that a field extension can often be assumed to have the separated basis property and that some type implications imply that the property can be lifted to a larger underlying field. In the subsequent section, we deduce strong consequences towards domination results from the separated basis property. Many results in earlier papers on domination used the assumption that the base C is maximal. Recall that a valued field is maximal (also called maximally complete or spherically complete) if it has no proper immediate extension. Here we show that this assumption can be replaced by the weaker assumption that there is a good separated basis over C.

Definition 2.1. Let M be a valued field extension of C. Let $V \subseteq M$ be a C-vector space. Let $m_1, \dots, m_k$ be elements of V, $\vec{m} = (m_1, \dots, m_k)$, and write $C \cdot \vec{m}$ for the C-vector subspace of V generated by $m_1, \dots, m_k$. We say that $\{m_1, \dots, m_k\}$ is a *separated basis over C* if for all $c_1, \dots, c_k$ in C,

$$v\left(\sum_{i=1}^{k} c_i m_i\right) = \min\{v(c_i m_i) : 1 \le i \le k\}$$

(and so, in particular, it forms a basis for $C \cdot \vec{m}$). We say that the separated basis is *good* if in addition for all $1 \le i, j \le k$, either $v(m_i) = v(m_j)$ or $v(m_i) - v(m_j) \notin \Gamma_C$. We say that V has the (*good*) *separated basis property over C* if every finite-dimensional C-subspace of V has a (good) separated basis.

By the next two lemmas, if the base C is either maximal or trivially valued, then any field extension has the good separated basis property.

Lemma 2.2 [HHM 2008, Proposition 12.1]. *Let C be a nontrivially valued maximal field and M a valued field extension. Then M has the good separated basis property over C.*

Lemma 2.3. *Let C be a trivially valued field, and M a nontrivially valued field extension. Then M has the good separated basis property over C.*

Proof. Since $v(c) = 0$ for every $c \in C$, the condition for being good is vacuous. To construct separated bases, let V be a finite-dimensional C-subspace of M, and proceed by induction on $\dim(V)$. If $\dim(V) = 1$ then any basis is automatically separated.

Assume the result is true for any ℓ-dimensional subspace, and let $\{m_1, \dots, m_\ell\}$ be a separated basis for $C \cdot \vec{m}$, the vector space that $\vec{m} = (m_1, \dots, m_\ell)$ generates over C. Assume without loss of generality that $v(m_1) \le v(m_2) \le \cdots \le v(m_\ell)$. Notice that, for all $m \in C \cdot \vec{m}$, $v(m) \in \{v(m_1), \dots, v(m_\ell)\}$. First suppose there is $m \in V \setminus C \cdot \vec{m}$ with $v(m) \notin \{v(m_1), \dots, v(m_\ell)\}$. Then $\{m_1, \dots, m_\ell, m\}$ is linearly independent and is separated. For suppose not. Then there are $c_1, \dots, c_{\ell+1}$ such that $v(c_{\ell+1} m) = v\big(\sum_{i=1}^{\ell} c_i m_i\big)$. Since $v(c_{\ell+1} m) = v(m)$ and $v\big(\sum_{i=1}^{\ell} c_i m_i\big) \in \{v(m_1), \dots, v(m_\ell)\}$, this contradicts the hypothesis on m.

Now suppose there is no such m. Let i_0 be the greatest $i \leq \ell$ for which there is $m \in V \setminus C \cdot \vec{m}$ with $v(m) = v(m_{i_0})$. We claim that $\{m_1, \ldots, m_\ell, m\}$ is a separated basis. Suppose not. Then there are some $c_1, \ldots, c_\ell, c_{\ell+1}$ for which the valuation of the sum is not given by the minimum. Write $I = \{i : v(m_i) = v(m_{i_0})\}$. In particular we must have (by induction)

$$v\left(\sum_{i \in I} c_i m_i + c_{\ell+1} m\right) > v(m_{i_0})$$

and $c_{\ell+1} \neq 0$. But then $\tilde{m} = \sum_{i \in I} c_i m_i + c_{\ell+1} m$ must have valuation that is not among the valuations of $m_1, \ldots, m_\ell$, or it must have valuation equal to $v(m_k)$ with $k > i_0$, which in either case contradicts our choice of m. □

Proposition 2.4. *Let C be a field and L a regular extension. Assume there is F a maximal immediate extension of C^{alg} such that $\mathrm{tp}_{\tilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\tilde{\mathcal{L}}}(L/F)$. Then L has the good separated basis property over C. Moreover, if C' is any algebraically closed field with $C \subseteq C' \subseteq F$, then the C'-vector space generated by L inside LF also has the good separated basis property over C'.*

Proof. If C is trivially valued, then the conclusion follows immediately from Lemma 2.3. So assume that C is not trivially valued.

The proof is by induction on the dimension of a finitely generated vector subspace of L over C. The base case is immediate, so assume for the induction hypothesis that $\ell_1, \ldots, l_{n+1}$ are linearly independent over C and that $\vec{\ell} = (\ell_1, \ldots, \ell_n)$ is a good separated basis not only for the space it generates over C but also for the space it generates over any algebraically closed C' with $C \subseteq C' \subseteq F$. By Lemma 1.20, $\ell_1, \ldots, l_{n+1}$ are linearly independent over F. As F is maximal (see the claim in the proof of [HHM 2008, Proposition 12.1]), there is a closest element of $F \cdot \vec{\ell}$ to ℓ_{n+1}; say

$$v\left(\sum_{i=1}^{n} b_i \ell_i - \ell_{n+1}\right) = \gamma$$

realizes this maximal valuation. Note that $\Gamma_F = \Gamma_{C^{\mathrm{alg}}}$ by choice of F and that $\Gamma(C) = \mathrm{dcl}_{\tilde{\mathcal{L}}}(C) \cap \Gamma = \Gamma(C^{\mathrm{alg}})$. Thus we may apply Lemma 1.17 to see that $\Gamma(LF) = \Gamma(L)$, and hence $\gamma \in \Gamma(L)$. In fact, applying Lemma 1.17 with L replaced by $L_0 = C(\ell_1, \ldots, \ell_{n+1})$ one sees that $\gamma \in \mathrm{dcl}_{\tilde{\mathcal{L}}}(C(\ell_1, \ldots, \ell_{n+1}))$.

Claim. *There is $b' \in C^{\mathrm{alg}} \cdot \vec{\ell}$ with $v(b' - \ell_{n+1}) = \gamma$.*

Proof of claim. Let $k = \mathrm{trdeg}(b_1, \ldots, b_n/C)$, assume that k is the minimum transcendence degree of any tuple in $\vec{d} \in F$ such that $v(\vec{d} \cdot \vec{\ell} - \ell_{n+1}) = \gamma$ and assume for contradiction that $k \geq 1$. Fix an algebraically closed $C' \subseteq C(b_1, \ldots, b_n)^{\mathrm{alg}}$ such that $\mathrm{trdeg}(C'(b_1, \ldots, b_n)/C') = 1$ and, without loss of generality, assume that $b_1 \notin C'$, that $b_2, \ldots, b_k \in C'$ are algebraically independent over C, and that

$\psi(b_1, \dots, b_k, x_{k+1}, \dots, x_n)$ is a formula which holds of $b_{k+1}, \dots, b_n$ and implies the algebraicity of $b_{k+1}, \dots, b_n$ over $C, b_1, \dots, b_k$.

Note that b_1 is also transcendental over $C'L$. For, since $\mathrm{tp}(L/C') \vdash \mathrm{tp}(L/F)$, we have that $\mathrm{tp}(LC'/C') \vdash \mathrm{tp}(LC'/F)$ and so $\mathrm{tp}(F/C') \vdash \mathrm{tp}(F/C'L)$ by Lemma 1.13. Hence if b_1 were algebraic over $C'L$, it would also be algebraic over C', which it is not.

Let $\varphi(x_1)$ be the formula

$$\exists x_{k+1} \dots \exists x_n \left(v\left(x_1\ell_1 + \sum_{i=2}^{k} b_i\ell_i + \sum_{i=k+1}^{n} x_i\ell_i - \ell_{n+1} \right) = \gamma \right) \\ \wedge \psi(x_1, b_2, \dots, b_k, x_{k+1}, \dots, x_n)$$

over $C'\ell_1 \dots \ell_{n+1}$. Since $\varphi(b_1)$ holds, and b_1 is not algebraic over $C'\vec{\ell}\ell_{n+1}$, we may assume that $\varphi(x_1)$ defines a finite union of $\mathrm{acl}_{\widetilde{\mathcal{L}}}(C'\vec{\ell}\ell_{n+1})$-definable swiss cheeses. Suppose for contradiction the swiss cheese containing b_1 does not intersect C'.

First note that all the points contained in it have the same $\widetilde{\mathcal{L}}$-type over C'. For suppose not. Then the outer ball of the swiss cheese contains a C'-definable closed ball of radius β. This closed ball contains infinitely many points of C' of distance β apart, which therefore cannot all be contained in the excluded balls of the swiss cheese, and hence at least one satisfies φ. It follows in particular that all extensions of C' generated by an element of this swiss cheese are isomorphic over C'.

There is a $d \in \mathrm{acl}_{\widetilde{\mathcal{L}}}(C'\vec{\ell}\ell_{n+1})$ realizing $\varphi(x_1)$, since this is a model. Because $\mathrm{tp}(d/C') = \mathrm{tp}(b_1/C')$ and $\mathrm{tp}(b_1/C') \vdash \mathrm{tp}(b_1/C'L)$, we have $\mathrm{tp}(d/C'L) = \mathrm{tp}(b_1/C'L)$. However, the extension $C'(d)$ cannot be isomorphic over $C'\vec{\ell}\ell_{n+1}$ to $C'(b_1)$, as b_1 is transcendental over $C'(\vec{\ell}\ell_{n+1})$.

This contradiction shows that there is $b'_1 \in C'$ realizing $\varphi(x_1)$ and hence also $b'_{k+1}, \dots, b'_n$ such that

$$v\left(b'_1\ell_1 + \sum_{i=2}^{k} b_i\ell_i + \sum_{i=k+1}^{n} b'_i\ell_i - \ell_{n+1} \right) = \gamma.$$

Since the formula $\psi(b'_1, b_2, \dots, b_k, x_{k+1}, \dots, x_n)$ holds of $b'_{k+1}, \dots, b'_n$, it follows that $b'_{k+1}, \dots, b'_n \in C'$. Thus $b'_1, b_2, \dots, b_k, b'_{k+1}, \dots, b'_n$ is a tuple in C' which witnesses the contradiction with the definition of k. □

Claim. *There is $b'' \in C \cdot \vec{\ell}$ with $v(b'' - \ell_{n+1}) = \gamma$.*

Proof of claim. We have $b' = \sum_{i=1}^{n} b'_i\ell_i \in C^{\mathrm{alg}} \cdot \vec{\ell}$ with $v(b' - \ell_{n+1}) = \gamma$. Let $\mathrm{Aut}(C^{\mathrm{alg}}/C)$ act on $b'_1, \dots, b'_n$ and let $b^1 = b', \dots, b^m$ be the conjugates of b' under this action. As $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C^{\mathrm{alg}})$ by assumption, and therefore $\mathrm{tp}(C^{\mathrm{alg}}/C) \vdash \mathrm{tp}(C^{\mathrm{alg}}/L)$, we have that for every $j < m$, $v(b^j - \ell_{n+1}) = \gamma$. Let

$b'' = \frac{1}{m}\sum_{j\le m} b^j$ (using the equicharacteristic 0 assumption). Then

$$v(b'' - \ell_{n+1}) = v\left(\frac{1}{m}\sum_{j\le m}(b^j - \ell_{n+1})\right) = \min_{j\le m}\{v(b^j - \ell_{n+1})\} = \gamma,$$

as the valuation cannot be greater than γ, by its definition. □

Now the argument is a straightforward calculation, as in [HHM 2008, Proposition 12.1]. We let $\ell'_{n+1} = \ell_{n+1} - b''$. Then $(\vec{\ell}, \ell'_{n+1})$ is a separated basis for the space it generates over F and hence also for the space generated over any subset of F, in particular for any C' with $C \subseteq C' \subseteq F$. Then, as in [HHM 2008, Lemma 12.2], the basis can be made into a good separated basis. □

As a corollary, we can show that the good separated basis property follows from stable domination. In the next section, we will prove that this characterizes stable domination.

Corollary 2.5. *Let a be a tuple of valued field elements, let C be a subfield of $\mathcal{U}$, and suppose that $L = \mathrm{dcl}_{\widetilde{\mathcal{L}}}(Ca)$ is a regular extension of C. If $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/C)$ is stably dominated then L has the good separated basis property over C.*

Proof. Working in $\widetilde{\mathcal{U}}$, let F be any immediate extension of C^{alg}. Because $\mathrm{St}_C(F) = \mathrm{St}_C(C^{\mathrm{alg}})$ and thus $\mathrm{St}_C(F) \subseteq \mathrm{acl}_{\widetilde{\mathcal{L}}}(\mathrm{St}_C(C))$, we have

$$\mathrm{St}_C(L) \mathop{\smile\!\!\!\!\!\!\!\perp}_C \mathrm{St}_C(F).$$

Because $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C)$ is stably dominated (and Proposition 1.15), we therefore have $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C\mathrm{St}_C(F)) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$. Clearly, $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C^{\mathrm{alg}}) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C\mathrm{St}_C(F))$ and as $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C^{\mathrm{alg}})$ by Lemma 1.19, we have $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$. If F is also maximal then we are in the situation of Proposition 2.4. □

The following lemma is stated as a claim in the proof of Proposition 12.11 of [HHM 2008] and the subsequent lemma is part of the statement of that proposition. However, in [HHM 2008], C is assumed to be maximal. We repeat the proofs here in order to clarify that the maximality of C is only used to obtain a separated basis.

Lemma 2.6. *Let L, M be valued fields with $C \subseteq L \cap M$ a valued subfield. Assume that $\Gamma_L \cap \Gamma_M = \Gamma_C$, k_L and k_M are linearly disjoint over k_C, and L has the good separated basis property over C. Choose $\{\ell_1, \dots, \ell_k\}$ a good separated basis for the subspace of L it generates over C. Then $\{\ell_1, \dots, \ell_k\}$ is still a good separated basis for the subspace of LM that it generates over M.*

Proof. Suppose, for a contradiction, that there are $m_1, \dots, m_k$ in M such that

$$v\left(\sum_{i=1}^{k} \ell_i m_i\right) > \min\{v(\ell_i m_i) : 1 \le i \le k\} = \gamma.$$

Let $I \subseteq \{1, \dots, k\}$ be the set of indices with $v(\ell_i m_i) = \gamma$ for $i \in I$. Note that $|I| > 1$ and for all i, j in I, $v(\ell_i) - v(\ell_j) = v(m_j) - v(m_i) \in \Gamma_L \cap \Gamma_M = \Gamma_C$. Thus $v(\ell_i) = v(\ell_j)$ as the basis is good. Fix $j \in I$ and write $I' = I \setminus \{j\}$. Now

$$v\Bigl(\sum_{i \in I} \ell_i m_i\Bigr) > \gamma \Longrightarrow v\Bigl(1 + \sum_{i \in I'} \frac{\ell_i m_i}{\ell_j m_j}\Bigr) > 0$$

and hence $\operatorname{res}\bigl(1+\sum_{i\in I'} \ell_i m_i/\ell_j m_j\bigr) = 0$. As $v(\ell_i/\ell_j) = v(m_i/m_j) = 0$, the residue map is a ring homomorphism, and hence

$$1 + \sum_{i \in I'} \operatorname{res}(\ell_i/\ell_j)\operatorname{res}(m_i/m_j) = 0.$$

As k_L, k_M are linearly disjoint over k_C, there must be $c_i \in C$ for $i \in I'$ with $\operatorname{res}(c_i)$ not all zero such that $\operatorname{res}(c_j) + \sum_{i \in I'} \operatorname{res}(\ell_i/\ell_j)\operatorname{res}(c_i) = 0$. Lifting back to the field gives $v\bigl(\sum_{i \in I'} \ell_i c_i\bigr) > v(\ell_j)$, which contradicts the assumption that $\{\ell_i : i \in I\}$ is separated over C. The basis is clearly good, as the value groups of L and M are disjoint over the value group of C. □

Lemma 2.6 gives the following purely algebraic statement.

Proposition 2.7. *Let L, M be valued fields with $C \subseteq L \cap M$ a valued subfield. Assume that $\Gamma_L \cap \Gamma_M = \Gamma_C$, that k_L and k_M are linearly disjoint over k_C and that L or M has the good separated basis property over C. Then L and M are linearly disjoint over C, Γ_{LM} is the group generated by Γ_L and Γ_M over Γ_C and k_{LM} is the field generated by k_L and k_M over k_C.*

Proof. Without loss of generality, L has the good separated basis property over C. To prove the linear disjointness, it suffices to show that any finite tuple $\ell_1, \dots, \ell_k$ from L which is linearly independent over C is also linearly independent over M (recall that we are working inside some ambient structure, so this statement makes sense). This follows from the conclusion of Lemma 2.6.

Now let x be in the ring generated by L and M over C. Then $x = \sum_{i=1}^{k} \ell_i m_i$ for some $\ell_i \in L$, $m_i \in M$ and we may assume that the ℓ_i form a good separated basis for the C-vector subspace of L that they generate. By Lemma 2.6 the tuple is also separated over M and hence $v(x) = v(\ell_j) + v(m_j)$ for some $j \in \{1, \dots, l\}$. Thus $\Gamma_{LM} = \Gamma_L \oplus_{\Gamma_C} \Gamma_M$. Suppose that $\operatorname{res}(x) \neq 0$. Let $I = \{i : v(\ell_i m_i) = 0\}$. Then $\operatorname{res}(x) = \operatorname{res}\bigl(\sum_{i \in I} \ell_i m_i\bigr) = \sum_{i \in I} \operatorname{res}(\ell_i m_i)$, and hence the residue field of k_{LM} is generated by k_L and k_M. □

3. Preliminary domination results

In this section, we show that a separated basis is strong enough to imply statements which are almost residue field domination results. The conclusion of Proposition 3.1 is not quite the statement of residue field domination for two reasons. Firstly, the type implication should be over the residue field of M, rather than the residue field of L. This is addressed in Corollary 3.2. Secondly, the type implication needs to be proved for subsets of any sort, not just the field sort. This is addressed in Section 4.

The first proposition shows that the good separated basis property is exactly what is needed in order to show type implication. The first part is a statement about $\widetilde{\mathcal{U}}$ and is Proposition 12.11 of [HHM 2008], except with the assumption of a good separated basis replacing the maximality of C. The further conclusion of this proposition is proved in [Ealy et al. 2019, Theorem 2.5] in the case of real closed valued fields. The proof given here is very similar, and illuminates the key properties to verify that the isomorphism of valued fields is actually an isomorphism of the full structure.

Proposition 3.1. *Let L, M be valued fields with $C \subseteq L \cap M$ a valued subfield. Assume that $\Gamma_L \cap \Gamma_M = \Gamma_C$, that k_L and k_M are linearly disjoint over k_C and that L or M has the good separated basis property over C. Let $\sigma : L \to L'$ be a valued field isomorphism which is the identity on C, Γ_L and k_L. Then σ extends by the identity on M to a valued field isomorphism from LM to $L'M$, and thus* $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/Ck_L\Gamma_L) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/M)$.

Suppose further that L and M are substructures of $\mathcal{U}$ and σ is an $\mathcal{L}$-isomorphism. Then σ is an isomorphism of RV_{LM} *to* $\mathrm{RV}_{L'M}$, *and thus* $\mathrm{tp}(L/Ck_L\Gamma_L) \vdash \mathrm{tp}(L/M)$.

Proof. By Proposition 2.7, L and M are linearly disjoint over C. Since $k'_L = k_L$, $\Gamma'_L = \Gamma_L$, and L' has the good separated basis property over C whenever L does, Proposition 2.7 also implies that L' and M are linearly disjoint over C. Hence σ extends to a field isomorphism on LM given by $\sigma\left(\sum \ell_i m_i\right) = \sum \sigma(\ell_i) m_i$ for any $\ell_i \in L$, $m_i \in M$.

To show that σ preserves the valuation on LM, choose x in the ring generated by L and M over C and write $x = \sum_{i=1}^{k} \ell_i m_i$. First suppose that L has the good separated basis property over C. We may assume that $\{\ell_1, \dots, \ell_k\}$ is separated over C and, as σ is a valued field isomorphism on L, this implies also that $\{\sigma(\ell_1), \dots, \sigma(\ell_k)\}$ is separated over C. Hence, by Lemma 2.6, both bases are separated over M. Then

$$v(x) = \min_{1 \le i \le k} \{v(\ell_i) + v(m_i)\} = \min_{1 \le i \le k} \{v(\sigma(\ell_i)) + v(m_i)\} = v(\sigma(x)),$$

as required. On the other hand, if we suppose that M has the good separated basis property over C, we may assume that $\{m_1, \dots, m_k\}$ is separated over C and hence, by Lemma 2.6, separated over L and L'. Then, as before,

$$v(x) = \min_{1 \le i \le k} \{v(\ell_i) + v(m_i)\} = \min_{1 \le i \le k} \{v(\sigma(\ell_i)) + v(m_i)\} = v(\sigma(x)),$$

as required.

Note that σ is the identity on k_L and k_M and hence by Proposition 2.7 on k_{LM}. Likewise, it is the identity on Γ_{LM}. Since $\sigma : LM \to L'M$ is a valued field isomorphism, it automatically preserves the group structure on RV_{LM}. Hence, to show that $\sigma : \mathrm{RV}_{LM} \to \mathrm{RV}_{L'M}$ is an isomorphism it suffices, by the quantifier elimination result in Proposition 1.3 and the fact that σ is the identity on Γ_{LM} and k_{LM}, to prove that the n-th power predicates are preserved; that is, $P_n(\mathrm{rv}(a)) \iff P_n(\sigma(\mathrm{rv}(a)))$. For each n, we have assumed there is a finite set of constants $\{\lambda\}$ which are representatives for the cosets of P_n. Of course, $\sigma(\lambda_L) = \lambda_{L'}$. Consider a coset representative ρ. Since for any $x, y \in RV$, whether or not xy is in the same coset as ρ depends only on the coset of x and the coset of y, we have for each ρ a finite set of pairs $\Lambda_{\rho,n} = \{(\lambda, \mu)\}$ such that

$$P_n(\rho^{-1}xy) \iff \bigvee_{(\lambda,\mu)\in\Lambda_{\rho,n}} P_n(\lambda x) \,\&\, P_n(\mu y).$$

Claim. *Suppose $a = \ell m$ for some $\ell \in L$, $m \in M$. Then for every n,*

$$P_n(\rho^{-1}\sigma(\mathrm{rv}(a))) \iff P_n(\rho^{-1}\,\mathrm{rv}(a)).$$

Proof of claim. We have

$$\begin{aligned}
P_n(\rho^{-1}\,\mathrm{rv}(a)) &\iff \bigvee_{(\lambda,\mu)\in\Lambda_{\rho,n}} P_n(\lambda\,\mathrm{rv}(\ell)) \,\&\, P_n(\mu\,\mathrm{rv}(m)) \\
&\iff \bigvee_{(\lambda,\mu)\in\Lambda_{\rho,n}} P_n(\sigma(\lambda\,\mathrm{rv}(\ell))) \,\&\, P_n(\sigma(\mu\,\mathrm{rv}(m))) \\
&\qquad\qquad\qquad \text{(as } \sigma|L \text{ is an isomorphism and } \sigma|M = \mathrm{Id}) \\
&\iff \bigvee_{(\lambda,\mu)\in\Lambda_{\rho,n}} P_n(\lambda\,\mathrm{rv}(\sigma(\ell))) \,\&\, P_n(\mu\,\mathrm{rv}(m)) \\
&\iff P_n(\rho^{-1}\sigma(\mathrm{rv}(a))). \qquad \square
\end{aligned}$$

Now let $a = \sum_{i=1}^{n} \ell_i m_i$ for some $n > 1$. By Proposition 2.7, $v(a)$ is in the group generated by Γ_L and Γ_M, so there are $\ell \in L$ and $m \in M$ with $v(a) = v(\ell m)$. Write $a = \ell m a_0$, where $v(a_0) = 0$, and note that $a_0 \in LM$. Then $\mathrm{rv}(a_0) = \mathrm{res}(a_0)$. As σ is the identity on k_{LM}, $\sigma(\mathrm{res}(a_0)) = \mathrm{res}(a_0)$, and therefore $\sigma(\mathrm{rv}(a_0)) = \mathrm{rv}(a_0)$. Thus $P_n(\sigma(\mathrm{rv}(a_0))) \iff P_n(\mathrm{rv}(a_0))$. Hence

$$\begin{aligned}
P_n(\mathrm{rv}(a)) &\iff \bigvee_{(\lambda,\mu)\in\Lambda_{1,n}} P_n(\lambda\,\mathrm{rv}(\ell m)) \,\&\, P_n(\mu\,\mathrm{rv}(a_0)) \\
&\iff \bigvee_{(\lambda,\mu)\in\Lambda_{1,n}} P_n(\lambda\sigma(\mathrm{rv}(\ell m))) \,\&\, P_n(\mu\sigma(\mathrm{rv}(a_0))) \\
&\qquad\qquad\qquad \text{(by the claim and the above)} \\
&\iff P_n(\sigma(\mathrm{rv}(a))). \qquad \square
\end{aligned}$$

As in [Ealy et al. 2019], it is helpful to state the following corollary, which means in particular that we can change the hypothesis on σ to assume that it fixes the value group and residue field of M instead of those of L.

Corollary 3.2. *Let L, M be substructures of $\mathcal{U}$ with $C \subseteq L \cap M$ a valued subfield. Assume that $\Gamma_L \cap \Gamma_M = \Gamma_C$, that k_L and k_M are linearly disjoint over k_C, and that L or M has the good separated basis property over C. Then* $\mathrm{tp}(L/C\Gamma_M k_M) \vdash \mathrm{tp}(L/M)$. *Similarly, if L and M are substructures of $\widetilde{\mathcal{U}}$ satisfying the same hypotheses, then* $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C\Gamma_M k_M) \vdash \mathrm{tp}\,\widetilde{\mathcal{L}}(L/M)$.

Proof. By Proposition 3.1, we have $\mathrm{tp}(L/C\Gamma_L k_L) \vdash \mathrm{tp}(L/M)$. Applying (v)⇒(ii) of Proposition 1.15, we obtain $\mathrm{tp}(L/C\Gamma_M k_M) \vdash \mathrm{tp}(L/M)$. □

Remark 3.3. If, in the preceding corollary, L could be taken from any sort, we would have proven the following: if $k(M)$ is a regular extension of $k(C)$, $\Gamma_M = \Gamma_C$, and M has the good separated basis property over C, then $\mathrm{tp}(M/C)$ is residue field dominated.

Corollary 3.2 often has implications for how forking behaves. When T is such that forking and dividing are the same, Corollary 3.2 describes circumstances in which forking in $\mathcal{U}$ can be reduced to forking in the residue field and value group, which is presumably easier to understand.

Corollary 3.4. *Assume that T implies that forking and dividing are the same over C, and assume further that $k(Ca)$ is a regular extension of k_C, $\Gamma(Ca)/\Gamma_C$ is torsion free, and either* $\mathrm{dcl}(Ca)$ *or* $\mathrm{dcl}(Cb)$ *has the good separated basis property over C. Then $a \perp\!\!\!\perp_C b$ if and only if $k(Ca)\Gamma(Ca) \perp\!\!\!\perp_C k(Cb)\Gamma(Cb)$.*

Proof. The proof is exactly that of Lemma 3.3(i) and Theorem 3.4(ii) of [Ealy et al. 2019], with the reference to Corollary 2.8 of that paper replaced by Corollary 3.2 of this one, and the use of elimination of imaginaries in the residue field replaced by strong stable embeddedness as in Remark 1.4. □

As a further corollary, we give a purely algebraic characterization of stable domination in ACVF (at least for a regular extension). We first note the following lemma.

Lemma 3.5. *Let C, L be valued fields with $C \subseteq L$ and suppose that L is henselian and an unramified regular extension of C. Then the following are equivalent:*

(1) *L has the good separated basis property over C.*

(2) $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$ *for some maximal immediate extension F of C^{alg}.*

(3) $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$ *for any maximal immediate extension F of C^{alg}.*

Proof. The implication (3) ⇒ (2) is clear and (2) ⇒ (1) is Proposition 2.4.

Let F be any maximal immediate extension of C^{alg} and assume that L has the good separated basis property over C. We apply Lemma 2.6 with C^{alg} replacing M. The lemma applies because L being henselian and regular implies that k_L is a regular extension of k_C: otherwise, there would be a polynomial with coefficients in k_C with a root in k_L, which would then lift to a polynomial over C with a root in L (as L is henselian and the residue characteristic is zero), contradicting the regularity of L over C. Applying Corollary 3.2, with LC^{alg} playing the role of L, C^{alg} playing the role of C, and F playing the role of M, we see that $\mathrm{tp}_{\widetilde{\mathcal{L}}}(LC^{\mathrm{alg}}/C^{\mathrm{alg}}) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(LC^{\mathrm{alg}}/F)$, and hence $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C^{\mathrm{alg}}) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$. Now apply Lemma 1.19 to obtain that $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$. □

Theorem 3.6. *Suppose that $\mathcal{U}$ is algebraically closed. Let $C \subset \mathcal{U}$ be a subfield, let a be a tuple of valued field elements, and let L be the definable closure of Ca in the valued field sort. Assume L is a regular extension of C. Then the following are equivalent*:

(i) $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/C)$ *is stably dominated.*

(ii) *L has the good separated basis property over C and L is an unramified extension of C.*

Proof. First assume (ii). Since L is definably closed, it is henselian. Thus we may apply Lemma 3.5 to see that $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/C) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/F)$ for some maximal extension F of C^{alg}. Applying Proposition 2.7, we see that $\Gamma_{LC^{\mathrm{alg}}} = \Gamma_{C^{\mathrm{alg}}}$. It follows that $\Gamma(LC^{\mathrm{alg}}) = \Gamma(C^{\mathrm{alg}})$, as both are equal to $\Gamma_{C^{\mathrm{alg}}}$. By [HHM 2008, Proposition 12.5], it follows that $\mathrm{tp}(a/C^{\mathrm{alg}})$ is orthogonal to Γ, which by Fact 1.6 is equivalent to being stably dominated. By Fact 1.9, $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/C)$ is stably dominated as $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/C^{\mathrm{alg}})$ is stably dominated.

The converse is handled by Corollary 2.5 along with the fact that stable domination implies orthogonality to the value group. □

RV-domination. As we recalled in Example 1.8, stable domination over the value group in an algebraically closed valued field [HHM 2008, Theorem 12.18] is implied by the assumptions that the base C is maximal, $k(L)$ is a regular extension of $k(C)$, and Γ_L/Γ_C is torsion free. We have already noted that this is not enough to get residue field domination over the value group. Here we introduce a notion of RV-domination, a property which does hold for the above example, and which in some ways feels closer to stable domination.

The analogue to the stable part of an algebraically closed valued field is here given by an infinite collection of definable subsets of RV, each of which is internal to the residue field. Let $M \supseteq C$ and $S \subset \Gamma$. Recall that $\mathrm{RV}_\gamma(M)$ is the fiber of the valuation map in $\mathrm{RV}(M)$ above γ, for $\gamma \in S$. Although this might seem to be very different from $\mathrm{St}_C(M)$, in fact, by [HHM 2008, Lemma 12.9], when C

and M are algebraically closed and S is definably closed, $\mathrm{acl}_{\widetilde{\mathcal{L}}}(\{\mathrm{RV}_\gamma(M)\}_{\gamma\in S})$ is essentially $\mathrm{St}_{CS}(M)$. Furthermore, [HHM 2008, Lemma 12.10] gives equivalent conditions for independence over $C\Gamma_L$ of $\mathrm{St}_{C\Gamma_L}(L)$ and $\mathrm{St}_{C\Gamma_L}(M)$. We take one of these equivalent conditions and use it as the definition of algebraic independence in RV.

Definition 3.7. Let L, M be subfields of $\mathcal{U}$ with $C \subseteq L \cap M$ a valued subfield. Assume that $\Gamma_L \subseteq \Gamma_M$ and Γ_L/Γ_C is torsion free. We say that $\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}$ is *algebraically independent* from $\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}$ over $C\Gamma_L$ if the following condition holds: for every sequence (a_i), (b_i) of elements of L, and (e_i) of elements of M such that

- **-:** $(v(a_i))$ is a $\mathbb{Q}$-basis for $\Gamma(L)$ over $\Gamma(C)$,
- **-:** $(\mathrm{res}(b_i))$ is a transcendence basis of k_L over k_C, and
- **-:** for all i, $v(a_i) = v(e_i)$,

the sequence $(\mathrm{res}(a_i/e_i), \mathrm{res}(b_j))$ is algebraically independent over $k(M)$.

Definition 3.8. Let $C \subseteq L$ be subfields of $\mathcal{U}$ such that Γ_L/Γ_C is torsion free. We say $\mathrm{tp}(L/C\Gamma_L)$ is RV-*dominated* if for any subfield $M \supseteq C$ such that $\Gamma_M \supset \Gamma_L$, if $\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}$ is algebraically independent from $\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}$ over $C\Gamma_L$ then

$$\mathrm{tp}(M/C\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}) \vdash \mathrm{tp}(M/L).$$

We note that this is not quite domination by RV, which is not a stable set in an algebraically closed valued field, but rather domination by a collection of k-internal sets. However, the more accurate name "RV_γ where γ ranges over Γ_L domination" is too unwieldy.

In order to prove a domination theorem, we first prove a result about extending isomorphisms. The following theorem was originally given in [HHM 2008, Proposition 12.15] in the case of algebraically closed valued fields, and then in [Ealy et al. 2019, Theorem 2.9] for real closed valued fields. The proof is somewhat subtle, and it is not completely obvious that the changes that are required for the current, more general, context carry through the machinery. For this reason, we repeat the proof in this paper, but postpone it to the Appendix.

Theorem 3.9. *Let L, M be subfields of $\widetilde{\mathcal{U}}$ with $C \subseteq L \cap M$ a valued subfield, $k(L)$ a regular extension of $k(C)$, and Γ_L/Γ_C torsion free. Assume that $\Gamma_L \subseteq \Gamma_M$, that $\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}$ is algebraically independent from $\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}$ over $C\Gamma_L$ and that L has the good separated basis property over C. Let σ be an automorphism of $\widetilde{\mathcal{U}}$ mapping L to L', which is the identity on C, Γ_L, and k_M. Then $\sigma|_L$ can be extended to a valued field isomorphism from LM to $L'M$ which is the identity on M. Furthermore, if σ is additionally the identity on RV_L, then σ may be extended to LM so that it is the identity on RV_{LM}.*

Theorem 3.10. *Let L, M be subfields of $\mathcal{U}$ with $C \subseteq L \cap M$ a valued subfield, $k(L)$ a regular extension of $k(C)$, $\Gamma_L \subseteq \Gamma_M$ and Γ_L/Γ_C torsion free. Assume that $\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}$ is algebraically independent from $\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}$ over $C\Gamma_L$ and that L has the good separated basis property over C. Let $\sigma : L \to L'$ be an $\mathcal{L}$-isomorphism which is the identity on C, $\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}$. Then σ can be extended by the identity on M to an automorphism of $\mathcal{U}$.*

Proof. We wish to show that $\mathrm{tp}(L/C\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L})$ implies $\mathrm{tp}(L/M)$. Observe that for each $\gamma \in \Gamma_L$, both $\mathrm{RV}_\gamma(L)$ and $\mathrm{RV}_\gamma(M)$ are nonempty. This (by Remark 1.4) allows us to apply (iv)⇒(i) of Proposition 1.15, and we see that it suffices to show that

$$\mathrm{tp}(L/C\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}) \vdash \mathrm{tp}(L/M).$$

The assumption that σ fixes $\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma(L)}$ implies that σ fixes k_M and Γ_L. By the above, we may assume that σ fixes $\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}$ as well. Thus we may apply Theorem 3.9 to get a valued field isomorphism $\sigma : LM \to L'M$ which is the identity on M and on RV_{LM}. In order to show that σ extends to an automorphism of $\mathcal{U}$, it suffices to show that it induces an isomorphism from the structure RV_{LM} to $\mathrm{RV}_{L'M}$, which is clear as the induced map is the identity. □

Theorem 3.11. *Let L be a subfield of $\mathcal{U}$ with $C \subseteq L$ a valued subfield. Assume that $k(L)$ is a regular extension of $k(C)$, Γ_L/Γ_C is torsion free and that L has the good separated basis property over C. Then* $\mathrm{tp}(L/C\Gamma_L)$ *is* RV*-dominated.*

Proof. Let M be a subfield of $\mathcal{U}$ as required in Definition 3.8. Theorem 3.10 gives us that $\mathrm{tp}(L/C\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}) \vdash \mathrm{tp}(L/M)$. As in the proof of Theorem 3.10, we may apply (i)⇔(ii) of Proposition 1.15 to obtain the type implication in the definition of RV-domination. □

4. The geometric sorts and domination

In the previous section, we worked within the field sort. However, our definition of residue field domination requires us to consider independent sets in any of the sorts. We thus need a mechanism to pull a hypothesis on an arbitrary geometric sort back to the field. This is given to us by the notion of a resolution.

The only sorts in $\mathcal{U}$, apart from the main sort, are RV and Γ. Of course, if one wanted to eliminate imaginaries, one would add more sorts including, but perhaps not limited to, the geometric sorts used to eliminate imaginaries in ACVF. The results in this section, proven as they are by carrying out the arguments of [HHM 2008] inside of $\mathcal{U}$, apply also to the geometric sorts. Thus for the remainder of this section, we take $\mathcal{U}$ to also refer to that portion of $\mathcal{U}^{\mathrm{eq}}$ consisting of the geometric sorts.

Definition 4.1. Let A be a subset of $\mathcal{U}$. We say that a set B in the field sort is a *resolution* of A if B is algebraically closed (in the sense of $\mathcal{L}$) in the field sort and $A \subseteq \operatorname{dcl}(B)$. The resolution is *prime* if B embeds over A into any other resolution.

In [HHM 2008, Theorem 11.14], the existence of prime resolutions is shown for algebraically closed valued fields. Thus, given $A \subset \mathcal{U} \subset \widetilde{\mathcal{U}}$, we have a resolution $B \subseteq \widetilde{\mathcal{U}}$, though it is not a priori clear that B would be contained in $\mathcal{U}$. Below, we give a careful analysis of the proof of the existence of resolutions, to see that the resolution can be constructed within $\mathcal{U}$. Since the proof involves checking that the arguments of various parts of Chapter 11 of [HHM 2008] never involve choosing something in $\widetilde{\mathcal{U}}$ that necessarily lies outside of $\mathcal{U}$, we follow the notation of [HHM 2008] as we walk the reader through this process. In particular, K refers to the field sort and R to the valuation ring.

Theorem 4.2. *Let $C \subseteq \mathcal{U}$ be a subfield, and let $e \in \mathcal{U}$ or more generally, in the geometric sorts of $\mathcal{U}$. Then Ce admits a resolution B with $k(B) = k(\operatorname{acl}(Ce))$ and $\Gamma(B) = \Gamma(Ce)$.*

Proof. We follow the construction in Chapter 11 of [HHM 2008], with the notation there. First, as in Theorem 11.14, we can assume that $e = (a, b)$, where $a \in B_n(K)/B_n(R)$ and $b \in B_m(K)/B_{m,m}(R)$. The next step is to replace e with an opaque layering of it (in the sense of ACVF). We need not concern ourselves here with the precise details of this, because we follow the construction in Lemmas 11.10 to 11.13 exactly. We need only check that the construction can be carried out in $\mathcal{U}$ and does not require elements of $\widetilde{\mathcal{U}} \setminus \mathcal{U}$. Through multiple applications of Lemma 11.10 and Corollary 11.11, $a = gB_n(R)$ is replaced successively by pairs $(h(H \cap F), \ell(N \cap F^h))$, where H, F are subgroups of $B_n(K)$, N is a normal subgroup of $B_n(K)$, $h \in H$, $\ell \in N$. Those subgroups are some of the G_i and H_i defined in Lemma 11.12, and are defined over $\mathbb{Z}$. The decomposition asserted in that lemma holds over any ring; in particular, it holds over our field $K(\mathcal{U})$. This shows that we can at each step take h and ℓ in $K(\mathcal{U})$. The same is true for b.

So we have replaced e by a sequence $\bar{a} = (a_0, \dots, a_{N-1})$ satisfying the conditions of Lemma 11.4 in the sense of ACVF and lying in $\mathcal{U}$. We therefore have $\operatorname{dcl}_{\widetilde{\mathcal{L}}}(C\bar{a}) = \operatorname{dcl}_{\widetilde{\mathcal{L}}}(Ce)$. Then we can find $C \subseteq D \subseteq K(\mathcal{U})$ such that $C\bar{a} \subset \operatorname{acl}_{\widetilde{\mathcal{L}}}(D)$ and D is atomic over $C\bar{a}$ (in $\widetilde{\mathcal{L}}$). This is by Lemma 11.4: all we do is take representatives of the equivalence relations defining the a_i (here $D = B_0 \cup C$ in the notation of Lemma 11.4). We can find such elements in $K(\mathcal{U})$ since $\bar{a}$ is in $\mathcal{U}$. Note that by the construction in Lemma 11.4, each representative is either in D or algebraic over D. In particular, each representative is contained in $\operatorname{acl}_{\widetilde{\mathcal{L}}}(D) \cap K(\mathcal{U})$.

Next, we want to expand D so that it remains atomic, but so that $C\bar{a}$ lies in the definable closure rather than the algebraic closure. We follow exactly the argument of Corollary 11.9, needing only to check that the construction does not leave $\mathcal{U}$. We know that $\bar{a}$ is in the definable closure of some $b \in \mathrm{acl}_{\widetilde{\mathcal{L}}}(D) \cap K(\mathcal{U})$ (namely the tuple of representatives). The orbit (in the sense of $\widetilde{\mathcal{U}}$) of b over $D\bar{a}$ is finite, and hence coded by some $b' \in K(\widetilde{\mathcal{U}})$. As b' is definable over a subset of $\mathcal{U}$, in particular b' is in $K(\mathcal{U})$. We thus have $b' \in \mathrm{dcl}_{\widetilde{\mathcal{L}}}(D\bar{a})$ with $\bar{a} \in \mathrm{dcl}_{\widetilde{\mathcal{L}}}(Db')$ and $\mathrm{tp}_{\widetilde{\mathcal{L}}}(Db'/C\bar{a})$ is isolated. (Note that our b is denoted e in Corollary 11.9, and our b' is denoted e'.)

From Corollary 11.16, we know that Ce admits a dcl-resolution B_0 such that $\mathrm{dcl}_{\widetilde{\mathcal{L}}}(B_0) \cap k = \mathrm{dcl}_{\widetilde{\mathcal{L}}}(Ce) \cap k$ and $\mathrm{dcl}_{\widetilde{\mathcal{L}}}(B_0) \cap \Gamma = \mathrm{dcl}_{\widetilde{\mathcal{L}}}(Ce) \cap \Gamma$. Referring to the proof of Corollary 11.16, we see that this dcl-resolution is the one obtained in Corollary 11.9. That is, $B_0 = Db'$, with D and b' as above. Let $B = \mathrm{acl}(Db') \cap K(\mathcal{U})$. To see that B is the required resolution, we just need to verify that $k(B) = k(\mathrm{acl}(Ce))$ and $\Gamma(B) = \Gamma(Ce)$.

First we show that $k(B_0) = k(Ce)$. It is clear that $k(B_0) \supseteq k(Ce)$, so take $d \in k(B_0)$, witnessed by φ. By quantifier elimination, φ is an $\mathcal{L}$-formula in the RV-sort and has the form $\varphi(x, \mathrm{rv}(t(Db')))$, where t is a term. Since there are no additional terms in $\mathcal{L}$ in the field sort, this is an $\widetilde{\mathcal{L}}$ term, and thus $\mathrm{rv}(t(Db')) \in \mathrm{dcl}_{\widetilde{\mathcal{L}}}(Db')$. From the proof that k is a stably embedded subset of RV, we may assume $\mathrm{rv}(t(Db'))$ is in k, and thus in $\mathrm{dcl}_{\widetilde{\mathcal{L}}}(B_0) \cap k = \mathrm{dcl}_{\widetilde{\mathcal{L}}}(Ce) \cap k$. Thus φ also witnesses that $d \in k(Ce)$.

Since it is clear that $k(B) \supseteq k(\mathrm{acl}(Ce))$, take $d_1 \in \mathrm{acl}(B_0) \cap k$. Suppose the conjugates of d_1 over B_0 are $d_1, \dots, d_n$. Then the set $\{d_1, \dots, d_n\}$ is in the definable closure of B_0 and, as fields code finite imaginaries, the set is coded by an element of $k(B_0) = k(Ce)$. Thus $d_1 \in \mathrm{acl}(Ce)$, as desired.

A similar argument shows that $\Gamma(B) = \Gamma(\mathrm{dcl}_{\widetilde{\mathcal{L}}}(Ce))$. □

By the following lemma, we see that proving a type implication for such a resolution is sufficient to give us the desired type implication that we need in the definition of residue field domination.

Lemma 4.3. *Fix a set of parameters C. Suppose that B is a resolution of Cb with $k(B) = k(\mathrm{acl}(Cb))$, and suppose that $\mathrm{tp}(a/Ck(B)) \vdash \mathrm{tp}(a/CB)$. Then*

$$\mathrm{tp}(a/Ck(Cb)) \vdash \mathrm{tp}(a/Cb).$$

Proof. Take $\varphi(x, b) \in \mathrm{tp}(a/Cb)$. Since $b \in \mathrm{dcl}(B)$, there is $\psi(x, d_1) \in \mathrm{tp}(a/Ck(B))$, which implies $\varphi(x, b)$. Consider the set $D = \{d_1, \dots, d_n\}$ of conjugates of d_1 over Cb. This set is definable over Cb, and thus so is $\bigvee_{d_i \in D} \psi(x, d_i)$. This latter formula is in $\mathrm{tp}(a/Ck(Cb))$ and implies $\varphi(x, b)$ as desired. □

The following lemma allows us to assume that elements are in the main sort when trying to prove domination results..

Lemma 4.4. *Fix* $\mathrm{tp}(a/C)$. *The following are equivalent*:

(i) *For any $b \in \mathcal{U}$, if $k(aC) \mathop{\smile\!\!\!\!\!\!\mid}^{\mathrm{alg}}_{k(C)} k(bC)$, then* $\mathrm{tp}(b/Ck(Ca)) \vdash \mathrm{tp}(b/Ca)$.

(ii) *For any b in the field sort of $\mathcal{U}$, if $k(aC) \downarrow^{\mathrm{alg}}_{k(C)} k(bC)$, then* $\mathrm{tp}(b/Ck(Ca)) \vdash \mathrm{tp}(b/Ca)$.

Proof. Clearly, (i) implies (ii). For the other direction, assume (ii) and choose $b \in \mathcal{U}$ such that $k(aC) \downarrow^{\mathrm{alg}}_{k(C)} k(bC)$. Choose a resolution B of Cb with $k(B) = k(\mathrm{acl}(Cb))$. As $k(aC) \downarrow^{\mathrm{alg}}_{k(C)} k(B)$, we conclude by (ii) that $\mathrm{tp}(B/Ck(Ca)) \vdash \mathrm{tp}(B/Ca)$ and thus by the equivalence of (i) and (ii) in Proposition 1.15 that $\mathrm{tp}(a/Ck(B)) \vdash \mathrm{tp}(a/CB)$. Then we may apply Lemma 4.3 to obtain $\mathrm{tp}(a/Ck(Cb)) \vdash \mathrm{tp}(a/Cb)$. We apply Proposition 1.15 again to obtain $\mathrm{tp}(b/Ck(Ca)) \vdash \mathrm{tp}(b/Ca)$. $\square$

As noted in Remark 3.3, Lemma 4.4 together with Corollary 3.2 gives us the following residue field domination result.

Theorem 4.5. *Let $C \subseteq \mathcal{U}$ be a subfield and let a be a (possibly infinite) tuple of field elements such that the field generated by Ca is an unramified extension of C with the good separated basis property over C, and such that $k(Ca)$ is a regular extension of $k(C)$. Then* $\mathrm{tp}(a/C)$ *is residue field dominated.*

Using Theorem 4.5 (or rather its component pieces: Corollary 3.2 and Lemma 4.4) we are able to push the above result a bit further and relate stable domination in the algebraically closed field to residue field domination in the henselian field. Recall that we write $C^+ = \mathrm{acl}(C) \cap \mathrm{dcl}(Ca)$.

Theorem 4.6. *Let $C \subseteq \mathcal{U}$ be a subfield and let $a \in \mathcal{U}$. Assume that $tp_{\widetilde{\mathcal{L}}}(a/C)$ is stably dominated. Then* $\mathrm{tp}(a/C^+)$ *is residue field dominated.*

Proof. First assume that a is a field element. By Fact 1.9, also $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/\mathrm{acl}(C))$ is stably dominated. Choose b with $k(\mathrm{acl}(C)a) \downarrow^{\mathrm{alg}}_{\mathrm{acl}(C)} k(\mathrm{acl}(C)b)$. By Lemma 4.4, we may assume that b is a field element. Let L be $\mathrm{dcl}(\mathrm{acl}(C)b)$ and let M be $\mathrm{dcl}(\mathrm{acl}(C)a)$. Since M is definably closed in $\mathcal{L}$ and thus also in $\widetilde{\mathcal{L}}$, it is a henselian valued field, and trivially M is a regular extension of $\mathrm{acl}(C)$, so we may use Corollary 2.5 to see that M has the good separated basis property over $\mathrm{acl}(C)$. Note that $\Gamma_M = \Gamma_{\mathrm{acl}(C)}$ by stable domination, so trivially $\Gamma_L \cap \Gamma_M = \Gamma_{\mathrm{acl}(C)}$. Since $k(\mathrm{acl}(C)a) \downarrow^{\mathrm{alg}}_{\mathrm{acl}(C)} k(\mathrm{acl}(C)b)$, Fact 1.18 implies k_L and k_M are linearly disjoint over $\mathrm{acl}(C)$. Thus Corollary 3.2 implies that

$$\mathrm{tp}(b/\mathrm{acl}(C)k(\mathrm{acl}(C)a)) \vdash \mathrm{tp}(b/\mathrm{acl}(C)a)$$

and hence $\mathrm{tp}(a/\mathrm{acl}(C))$ is residue field dominated. By Proposition 1.12, $\mathrm{tp}(a/C^+)$ is residue field dominated.

Now let a be in any of the sorts. By Facts 1.9 and 1.6, $\mathrm{tp}_{\widetilde{\mathcal{L}}}(a/\mathrm{acl}(C))$ is orthogonal to Γ. By [HHM 2008, Lemma 10.14], there is a resolution B of $\mathrm{acl}(C)a$ such that $\mathrm{tp}(B/C)$ is orthogonal to Γ. On the other hand, we know by Theorem 4.2 and [HHM 2008, Theorem 11.14], that $\mathrm{acl}(C)a$ has a prime resolution A that only adds algebraic elements to $k(Ca)$ and lies in $\mathcal{U}$. By primality, A embeds into B and hence its $\widetilde{\mathcal{L}}$-type is also orthogonal to Γ, so also stably dominated. By Theorem 4.6, $\mathrm{tp}(A/C^+)$ is residue field dominated. Consider any $b \in \mathcal{U}$ such that $k(C^+b) \mathop{\downarrow}^{\mathrm{alg}}_{C^+} k(C^+a)$. Since $k(A) = \mathrm{acl}(k(C^+a))$, we have $k(C^+b) \mathop{\downarrow}^{\mathrm{alg}}_{C^+} k(A)$. By residue field domination for $\mathrm{tp}(A/C^+)$, we have $\mathrm{tp}(b/C^+k(A)) \vdash \mathrm{tp}(b/C^+A)$. Now apply Lemma 4.3 to see that $\mathrm{tp}(b/C^+k(C^+a)) \vdash \mathrm{tp}(b/C^+A)$. □

Appendix: Proof of Theorem 3.9

This proof is essentially the same as that given in [HHM 2008, Proposition 12.15] in the case of algebraically closed valued fields, and then in [Ealy et al. 2019, Theorem 2.9] for real closed valued fields. In the other two papers, the fields L, M, and C are assumed to be algebraically (respectively real) closed. We show that this hypothesis is not really needed. We also show that the prior assumption that C is maximal can be replaced with the good separated basis property for L over C. Furthermore, we prove the additional conclusion that if σ is the identity on RV_L as well, then σ extends by the identity to all of RV_{LM}.

Theorem 3.9. *Let L, M be subfields of $\widetilde{\mathcal{U}}$ with $C \subseteq L \cap M$ a valued subfield, $k(L)$ a regular extension of $k(C)$, and Γ_L/Γ_C torsion free. Assume that $\Gamma_L \subseteq \Gamma_M$, that $\{\mathrm{RV}_\gamma(L)\}_{\gamma\in\Gamma_L}$ is algebraically independent from $\{\mathrm{RV}_\gamma(M)\}_{\gamma\in\Gamma_L}$ over $C\Gamma_L$, and that L has the good separated basis property over C. Let σ be an automorphism of $\widetilde{\mathcal{U}}$ mapping L to L' which is the identity on C, Γ_L, and k_M. Then $\sigma|_L$ can be extended to a valued field isomorphism from LM to $L'M$ which is the identity on M. Furthermore, if σ is additionally the identity on RV_L, then σ may be extended to LM so that it is the identity on RV_{LM}.*

Proof. In outline, we begin by perturbing the valuation to a finer one, v', which satisfies the hypothesis that $\Gamma_{(L,v')} \cap \Gamma_{(M,v')} = \Gamma_{(C,v')}$. We can then apply Proposition 3.1 to extend $\sigma|_L$ to a v'-valued field isomorphism from LM to $L'M$ which extends the identity on M. An analysis of the construction shows that this is also a v-valued field isomorphism. Finally, we use the separated basis hypothesis to show that σ is also an isomorphism on RV_{LM}.

The first statement to be proved can be rephrased as saying

$$\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/Ck_M\Gamma_L) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/M).$$

To prove this, we claim that it suffices to prove $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/Ck_M\Gamma_L\Gamma_M) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/M)$. For, by Lemma 1.14, with Ck_M replacing C, and Γ replacing $\mathcal{S}$, we know that $\mathrm{tp}_{\widetilde{\mathcal{L}}}(L/Ck_M\Gamma(k_ML)) \vdash \mathrm{tp}_{\widetilde{\mathcal{L}}}(L/Ck_M\Gamma(k_ML)\Gamma(M))$. Thus, we just need to verify that $\Gamma(k_ML) = \Gamma(L) = \Gamma_L$. This follows by orthogonality of the value group and residue field. Thus we may assume that σ fixes Γ_M as well.

Choose $a_1, \dots, a_r$ from L and $e_1, \dots, e_r$ from M such that, for each $1 \le i \le r$, $v(a_i) = v(e_i)$ and $\{v(a_i)\}$ forms a $\mathbb{Q}$-basis for Γ_L modulo Γ_C. Choose $b_1, \dots, b_s$

from L such that $\{\mathrm{res}(b_1), \dots, \mathrm{res}(b_s)\}$ is a transcendence basis for k_L over k_C. By Definition 3.7, the elements

$$\mathrm{res}(a_1/e_1), \dots, \mathrm{res}(a_r/e_r), \mathrm{res}(b_1), \dots, \mathrm{res}(b_s)$$

are algebraically independent over k_M. For $0 \le j \le r$, let

$$R^{(j)} = \mathrm{acl}(k_M, \mathrm{res}(a_1/e_1), \dots, \mathrm{res}(a_j/e_j), \mathrm{res}(b_1), \dots, \mathrm{res}(b_s)) \cap k_{LM}.$$

In particular,

$$\begin{aligned} R^{(0)} &= \mathrm{acl}(k_M, \mathrm{res}(b_1), \dots, \mathrm{res}(b_s)) \cap k_{LM} = \mathrm{acl}(k_M, k_L) \cap k_{LM}, \\ R^{(r)} &= \mathrm{acl}(k_M, \mathrm{res}(a_1/e_1), \dots, \mathrm{res}(a_r/e_r), k_L) \cap k_{LM}. \end{aligned}$$

For each $0 \le j \le r-1$, choose a place $p^{(j)} : R^{(j+1)} \to R^{(j)}$ fixing $R^{(j)}$ and such that $p^{(j)}(\mathrm{res}(a_{j+1}/e_{j+1})) = 0$, which is possible by the algebraic independence of $\mathrm{res}(a_1/e_1), \dots, \mathrm{res}(a_r/e_r)$ over k_M. Also choose a place $p^* : k_{LM} \to R^{(r)}$ fixing $R^{(r)}$. (Later we will show that $k_{LM} = R^{(r)}$ and thus p^* will be seen to be the identity.) Write $p_v : LM \to k_{LM}$ for the place corresponding to our given valuation v. Define $p_{v'} : LM \to R^{(0)}$ to be the composition

$$p_{v'} = p^{(0)} \circ \dots \circ p^{(r-1)} \circ p^* \circ p_v.$$

Let v' be a valuation associated to the place $p_{v'}$. Notice that all the places $p^{(j)}$ and p^* are the identity on k_M, so we may identify (M, v) and (M, v'), including identifying the value groups Γ_M and $\Gamma_{(M,v')}$. Similarly, the places are all the identity on k_L, so the value groups Γ_L and $\Gamma_{(L,v')}$ are isomorphic, but we shall see that we cannot simultaneously identify Γ_M with $\Gamma_{(M,v')}$ and Γ_L with $\Gamma_{(L,v')}$.

We now have two valuations v and v' on LM. If $x \in M \subseteq LM$, then $v(x) = v'(x)$, and if $x, y \in L \subseteq LM$ then $v(x) \le v(y)$ implies $v'(x) \le v'(y)$. Furthermore, the construction has ensured that for any $x \in M$ with $v(x) > 0$, and any w such that $\mathrm{res}(w)$ is a nonzero element of k_{LM} mapped to zero by p^*,

$$0 < v'(a_1/e_1) \ll \dots \ll v'(a_r/e_r) \ll v(w) \ll v'(x),$$

where $\gamma \ll \delta$ means that $n\gamma < \delta$ for any $n \in \mathbb{N}$ (and hence $\Gamma_{(L,v')} \neq \Gamma_L$). Let Δ be the subgroup of $\Gamma_{(LM,v')}$ generated by $v'(a_1/e_1), \dots, v'(a_r/e_r)$ together with $v(w)$ for all such w. Then Δ is a convex subgroup of $\Gamma_{(LM,v')}$ and $\Gamma_{(LM,v')} = \Delta \oplus \Gamma_{LM}$, where the right-hand group is ordered lexicographically. (See, e.g., Theorem 15, Theorem 17, and the associated discussion in Chapter VI of [Zariski and Samuel 1975]).

To see that $\Gamma_{(L,v')} \cap \Gamma_{(M,v')} = \Gamma_{(C,v')}$, let $m \in M$ and $\ell \in L$ be such that $v'(m) = v'(\ell)$. Set $v'(a_i/e_i) = \delta_i$ and $v'(e_i) = \epsilon_i$. As $(v(a_i))$ generates Γ_L over Γ_C, and Γ_L and $\Gamma_{(L,v')}$ are isomorphic,

$$v'(\ell) = \sum_{i=1}^{r} p_i v'(a_i) + \gamma = \sum_{i=1}^{r} p_i \delta_i + \sum_{i=1}^{r} p_i \epsilon_i + \gamma,$$

where $p_i \in \mathbb{Q}$ and $\gamma \in \Gamma_C$. The set

$$\{\delta_1, \dots, \delta_r, \epsilon_1, \dots, \epsilon_r\}$$

is algebraically independent over Γ_C since $\Gamma_{(LM,v')} = \Delta \oplus \Gamma_{LM}$. Next, note that since $v'(e_i) = v(e_i)$, $\{v'(e_i)\}$ forms a $\mathbb{Q}$-basis of $\Gamma_L \subseteq \Gamma_M = \Gamma_{(M,v')}$ over Γ_C. Let $\mu_1, \dots, \mu_t$ be such that $\{\epsilon_i\} \cup \{\mu_j\}$ forms a $\mathbb{Q}$-basis of Γ_M over Γ_C. Then

$$v'(m) = \sum_{i=1}^{r} p_i' \epsilon_i + \sum_{i=1}^{t} q_i \mu_i + \gamma',$$

where $q_i \in \mathbb{Q}$ and $\gamma' \in \Gamma_C$. It follows that each $p_i = p_i' = 0$ and each $q_i = 0$, hence $v'(\ell) = v'(m) \in \Gamma_C$.

Next we must check that $k_{(L,v')}$ and $k_{(M,v')}$ are linearly disjoint. Our definition of RV-independence implies that k_L and k_M are independent over k_C, and using that k_L is a regular extension of k_C and Fact 1.18 we obtain that k_L and k_M are linearly disjoint over k_C. As already observed, the place

$$p^{(0)} \circ \cdots \circ p^{(r-1)} \circ p^* : k_{LM} \to \operatorname{acl}(k_M, k_L) \cap k_{LM}$$

is the identity on k_M and k_L. Thus this place is also the identity on their compositum, and $k_L k_M = k_{(L,v')} k_{(M,v')}$. Thus k_L and k_M being linearly disjoint over k_C implies linear disjointness of $k_{(L,v')}$ and $k_{(M,v')}$ over $k_{(C,v')}$.

Hence we can apply Corollary 3.2 to deduce that the isomorphism $\sigma|_L$ extends to a valued field isomorphism from (LM, v') to $(L'M, v')$ which is the identity on M. As v' is a refinement of v, σ is also an isomorphism of (LM, v).

Moreover, by Proposition 2.7, we know that $\Gamma_{(LM,v')}$ is the sum of $\Gamma_{(L,v')}$ and $\Gamma_{(M,v')}$, and $k_{(LM,v')} = k_{(L,v')} k_{(M,v')}$. Since $\Gamma_{(LM,v')}$ is also $\Delta \oplus \Gamma_{LM}$, we see both that Δ must be generated by $\delta_1, \dots, \delta_r$ and that $\Gamma_{LM} = \Gamma_M$. Since Δ is generated by $\delta_1, \dots, \delta_r$, in particular this means that there is no w such that $\operatorname{res}(w)$ is a nonzero element mapped to zero by p^*. This implies that p^* is the identity, and that $k_{LM} = \operatorname{acl}(k_M, \operatorname{res}(a_1/e_1), \dots, \operatorname{res}(a_r/e_r), k_L) \cap k_{LM}$.

It remains to show that if σ is the identity on RV_L, then it is also the identity on RV_{LM}. Take an element of LM, say $\left(\sum_{i<n} \ell_i m_i\right) / \left(\sum_{j<n} \ell_j m_i'\right)$. By the hypothesis, we may assume that the $\{\ell_i\}$ forms a good separated basis over C with respect to v for the subspace it generates, and also with respect to v', since (L, v) and (L, v') are isomorphic. By Lemma 2.6, this basis is still separated over M with

respect to v'. Hence, it is even separated over M with respect to v, as the following calculation shows:

$$v\Big(\sum_{i<n} m_i\ell_i\Big) = v'\Big(\sum_{i<n} m_i\ell_i\Big)/\Delta = (\min_{i<n}\{v'(m_i\ell_i)\})/\Delta$$
$$= \min_{i<n}\{v'(m_i\ell_i)/\Delta\} = \min_{i<n}\{v(m_i\ell_i)\}.$$

Since the basis is separated, we can calculate the rv of an element of RV_{LM} as below. Let I be the set of indices when $v(m_i\ell_i)$ attains its minimum. Then

$$\mathrm{rv}\Big(\sum_{i<n} m_i\ell_i\Big) = \mathrm{rv}\Big(\sum_{i\in I} m_i\ell_i\Big) = \sum_{i\in I}\mathrm{rv}(m_i\ell_i) = \sum_{i\in I}\mathrm{rv}(m_i)\,\mathrm{rv}(\ell_i).$$

As σ fixes RV_L and M, we see that σ fixes rv of any element of the form $\sum_{i<n} m_i\ell_i$. Hence σ fixes rv of any element which is a quotient of such elements, i.e., any element of LM. □

Acknowledgements

We thank Mariana Vicaria for comments on previous versions of this paper and in particular for pointing out a mistake in the statement of Theorem 3.10. We also thank the referee for very carefully reading the paper and compiling an exhaustive list of comments which resulted in great improvements to the exposition.

References

[Chernikov and Simon 2019] A. Chernikov and P. Simon, "Henselian valued fields and inp-minimality", *J. Symb. Log.* **84**:4 (2019), 1510–1526. MR Zbl

[Cluckers and Loeser 2007] R. Cluckers and F. Loeser, "b-minimality", *J. Math. Log.* **7**:2 (2007), 195–227. MR Zbl

[Cubides Kovacsics et al. 2021] P. Cubides Kovacsics, M. Hils, and J. Ye, "Beautiful pairs", preprint, 2021. arXiv 2112.00651

[Ealy et al. 2019] C. Ealy, D. Haskell, and J. Maříková, "Residue field domination in real closed valued fields", *Notre Dame J. Form. Log.* **60**:3 (2019), 333–351. MR Zbl

[Fehm and Jahnke 2016] A. Fehm and F. Jahnke, "Fields with almost small absolute Galois group", *Israel J. Math.* **214**:1 (2016), 193–207. MR Zbl

[HHM 2006] D. Haskell, E. Hrushovski, and D. Macpherson, "Definable sets in algebraically closed valued fields: elimination of imaginaries", *J. Reine Angew. Math.* **597** (2006), 175–236. MR Zbl

[HHM 2008] D. Haskell, E. Hrushovski, and D. Macpherson, *Stable domination and independence in algebraically closed valued fields*, Lect. Notes in Logic **30**, Cambridge University Press, 2008. MR Zbl

[Hrushovski and Loeser 2016] E. Hrushovski and F. Loeser, *Non-archimedean tame topology and stably dominated types*, Annals of Math. Studies **192**, Princeton University Press, 2016. MR Zbl

[Hrushovski and Rideau-Kikuchi 2019] E. Hrushovski and S. Rideau-Kikuchi, "Valued fields, metastable groups", *Selecta Math.* (*N.S.*) **25**:3 (2019), paper no. 47. MR Zbl

[Lang 2002] S. Lang, *Algebra*, 3rd ed., Grad. Texts in Math. **211**, Springer, 2002. MR Zbl

[Pas 1989] J. Pas, "Uniform p-adic cell decomposition and local zeta functions", *J. Reine Angew. Math.* **399** (1989), 137–172. MR Zbl

[Rideau-Kikuchi 2017] S. Rideau-Kikuchi, "Some properties of analytic difference valued fields", *J. Inst. Math. Jussieu* **16**:3 (2017), 447–499. MR Zbl

[Scanlon 2003] T. Scanlon, "Quantifier elimination for the relative Frobenius", pp. 323–352 in *Valuation theory and its applications* (Saskatoon, SK, 1999), vol. II, edited by F.-V. Kuhlmann et al., Fields Inst. Commun. **33**, Amer. Math. Soc., Providence, RI, 2003. MR Zbl

[Vicaria 2021] M. Vicaria, "Residue field domination in henselian valued fields of equicharacteristic zero", preprint, 2021. arXiv 2109.08243

[Zariski and Samuel 1975] O. Zariski and P. Samuel, *Commutative algebra*, vol. II, Grad. Texts in Math. **29**, Springer, 1975. MR Zbl

Received 28 Feb 2022. Revised 26 May 2023.

CLIFTON EALY:

cf-ealy@wiu.edu
Department of Mathematics and Philosophy, Western Illinois University, Macomb, IL, United States

DEIRDRE HASKELL:

haskell@math.mcmaster.ca
Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada

PIERRE SIMON:

simon@math.berkeley.edu
Department of Mathematics, University of California, Berkeley, CA, United States

msp
Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.285

Star sorts, Lelek fans, and the reconstruction of non-$\aleph_0$-categorical theories in continuous logic

Itaï Ben Yaacov

We prove a reconstruction theorem valid for arbitrary theories in continuous (or classical) logic in a countable language, that is to say that we provide a complete bi-interpretation invariant for such theories, taking the form of an open Polish topological groupoid.

More explicitly, for every such theory T we construct a groupoid $\boldsymbol{G}^*(T)$ that only depends on the bi-interpretation class of T, and conversely, we reconstruct from $\boldsymbol{G}^*(T)$ a theory that is bi-interpretable with T. The basis of $\boldsymbol{G}^*(T)$ (namely, the set of objects, when viewed as a category) is always homeomorphic to the Lelek fan.

We break the construction of the invariant into two steps. In the second step we construct a groupoid from any sort of codes for models, while in the first step such a sort is constructed. This allows us to place our result in a common framework with previously established ones, which only differ by their different choice of sort of codes.

Introduction

This paper deals with what we have come to refer to as *reconstruction theorems*. By this we mean a procedure that associates to a theory T (possibly under some

---

Author supported by ANR project AGRUME (ANR-17-CE40-0026).
MSC2020: 03C15, 03C30, 03C95, 22A22.
Keywords: continuous logic, theory, interpretation, bi-interpretation, sort, topological groupoid, reconstruction.

hypotheses) a topological group-like object that is a complete bi-interpretation invariant for T. In other words, if T' is bi-interpretable with T, then we associate to it the same object (up to an appropriate notion of isomorphism), and conversely, the isomorphism class of this object determines the bi-interpretation class of T.

The best-known result of this kind is due to Coquand, and appears in [Ahlbrandt and Ziegler 1986]. It states that if T is an $\aleph_0$-categorical theory (in a countable language), then the topological group $G(T) = \mathrm{Aut}(M)$, where M is the unique countable model, is such an invariant. This was originally proved for theories in classical (Boolean-valued) logic, and subsequently extended by Kaïchouh and the author [Ben Yaacov and Kaïchouh 2016] to continuous (real-valued) logic.

In [Ben Yaacov 2022] we proposed a reconstruction result that also covers some non-$\aleph_0$-categorical theories, using a topological groupoid (rather than a group) as invariant. The result was presented in two forms, first for classical logic and then for the more general continuous logic. This was not done for the sake of presentation (doing the more familiar case first), but because of a fundamental difference between the two cases. In classical logic, we have a straightforward construction of a sort of "codes of models" (more about this later). In continuous logic, on the other hand, no such construction exists in general, and we were reduced to assuming that such a sort (satisfying appropriate axioms) existed, and was given to us. Worse still, we gave an example of a theory for which no such sort existed, and consequently, for which our reconstruction theorem was inapplicable.

In the present paper we seek to remedy this deficiency, proposing a reconstruction theorem that holds for all theories (in a countable language). This time, we work exclusively in continuous logic, keeping in mind that this contains classical logic as a special case.

In Section 1 we provide a few reminders regarding continuous logic in general, and interpretable sorts in particular. We (re)define the notions of interpretation and bi-interpretation, in a manner that is particularly appropriate for the use we shall make of them, and that avoids the rather tedious notions of interpretation schemes.

In Section 2 we discuss various ways in which one sort E can be "coded" in another sort D, both uniform (e.g., E is interpretable in D) and nonuniform (e.g., each $a \in E$ is in the definable closure of some $b \in D$). We define a *coding sort* D as a sort which codes models. Every sort is coded in a coding sort in a nonuniform fashion, and therefore in a uniform fashion as well.

In Section 3 we associate to a coding sort D a topological groupoid $\boldsymbol{G}_D(T)$, from which a theory T_{2D}, bi-interpretable with T, can be recovered. In particular, $\boldsymbol{G}_D(T)$ determines the bi-interpretation class of T. If, in addition, D only depends on the bi-interpretation class of T, then so does $\boldsymbol{G}_D(T)$, in which case it is a complete bi-interpretation invariant. We point out, rather briefly, how previous reconstruction theorems fit in this general setting.

In Sections 4 and 5 we define *star spaces* and *star sorts*. These, by their very nature, require us to work in continuous (rather than classical) logic. In particular, we define a notion of a *universal star sort*, and show that if it exists, then it is unique up to definable bijection, and only depends on the bi-interpretation class of T.

In Section 6 we use the star sort formalism to give a construction that is analogous to, though not a direct generalisation of, the construction of the coding sort for classical theories in [Ben Yaacov 2022]. We then prove that the resulting sort is a universal star sort, so one always exists. Moreover, the construction is independent of the theory: we simply construct, for any countable language $\mathcal{L}$, a star sort D^* that is universal in any $\mathcal{L}$-theory, complete or incomplete.

We conclude in Section 7, showing that the universal star sort must be a coding sort, whence our most general reconstruction theorem: in a countable language, the groupoid $\boldsymbol{G}_{D^*}(T)$ is a complete bi-interpretation invariant for T. We also show that the type-space of the sort D^*, relative to any complete theory T, is the Lelek fan L. Finally, in case T does fall into one of the cases covered by previous results, we show that our last result can be viewed as some kind of generalisation. More precisely, using the Lelek fan, we can recover the coding sort D^*, and therefore the corresponding groupoid $\boldsymbol{G}_{D^*}(T)$, from those given by the earlier results.

1. Sorts and interpretations

As said in the introduction, we work exclusively in continuous first order logic, and assume that the reader is familiar with it. For a general exposition, see [Ben Yaacov and Usvyatsov 2010; Ben Yaacov et al. 2008]. We allow formulas to take truth values in arbitrary compact subsets of $\mathbf{R}$, so connectives are arbitrary continuous functions from $\mathbf{R}^n$ to $\mathbf{R}$. For a countable family of connectives, it suffices to take all rational constants, addition and multiplication, to which we add the absolute value operation. Closing these under composition yields a (countable) family of functions that is dense among all continuous functions on each compact subset of $\mathbf{R}^n$.

Notation 1.1. Using the absolute value operation we may define maximum and minimum directly (i.e., without passing to a limit). We use infix notation $\vee$ and $\wedge$ for those. We also write $t \dot{-} s$ for the *truncated subtraction* $(t - s) \vee 0$.

We allow the language to be many-sorted. Some of the time we also require the language to be countable, which means in particular that the set of sorts is countable, although this is not a requirement for the present section.

We are going to talk quite a bit about sorts and interpretations, so let us begin with a few reminders. By a *sort* we mean an interpretable sort in the sense of continuous logic, as discussed, for example, in [Ben Yaacov and Kaïchouh 2016; Ben Yaacov 2022]. Sorts are obtained by closing the family of basic sorts (namely, sorts named in the language) by

- adding the constant sort $\{0, 1\}$ (so it is always implicitly interpretable),
- countable product,
- quotient by a definable pseudodistance (in a model that is not saturated, this may also require a passage to the completion), and
- nonempty definable subset.

We follow the convention that a natural number $n \in \mathbf{N}$ is coded by the set $\{0, \dots, n-1\}$, so $\{0, 1\}$ may sometimes be denoted by 2 (this is especially true of its powers: the Cantor space is $2^{\mathbf{N}}$).

Throughout, by *definable* we mean definable by a formula, without parameters (unless parameters are given explicitly). Any function $\{0, 1\} \to \mathbf{R}$ is a formula on the sort $\{0, 1\}$. Formulas on a finite product of sorts are constructed in the usual way, using function and predicate symbols, connectives and quantifiers, and closing the lot under uniform limits. In particular, if $\varphi_i(x)$ are formulas on a sort D for $i < 2^n$, then $\varphi(i, x) = \varphi_i(x)$ is a formula on $2^n \times D$. Formulas on an infinite product of sorts consist of all formulas on finite subproducts (extended to the whole product through the addition of dummy variables), as well as all uniform limits of such (where the subproducts through which they factor may vary). If $\bar{d}$ is a definable pseudodistance on a sort D (defined by a formula on $D \times D$), then formulas on the quotient $(D, \bar{d})$ are formulas on D that are uniformly continuous with respect to $\bar{d}$, and similarly for formulas on a product of several quotient sorts.

Finally, we recall that a definable subset of a sort D is a subset $E \subseteq D$, the distance to which is definable (this is significantly more involved than the notion of a definable subset in classical logic). Equivalently, if for every formula $\varphi(x, y)$, where x is a variable in D and y is a tuple of variables in arbitrary sorts, the predicate $\sup_{x \in E} \varphi(x, y)$ is definable by a formula $\psi(y)$. Formulas on a product of definable subsets of sorts are restrictions of formulas on the corresponding product of ambient sorts.

Notice that every compact metric space is a quotient space of $2^{\mathbf{N}}$ by a continuous pseudodistance, and therefore a sort, on which the formulas are the continuous functions. Conversely, we could have chosen any nontrivial compact metric space as a basic constant sort in place of $\{0, 1\}$ (the other obvious candidate being $[0, 1]$), and realise $\{0, 1\}$ as any two-point set therein.

Remark 1.2. An obvious, yet crucial remark, is that if $\varphi(x, y)$ is an arbitrary formula on $E \times D$, then

$$d_\varphi(y, y') = \sup_{x \in E} |\varphi(x, y) - \varphi(x, y')|$$

defines a pseudodistance on D. In addition, if $D = E$, and φ happens to define a pseudodistance on D, then it agrees with d_φ.

This has numerous useful consequences, let us state two of them explicitly. First of all, one may be bothered by the fact that a formula $\varphi(x, y)$ defining a pseudodistance on a sort D may depend on the structure(s) under consideration. However, we may restrict the "quotient by a pseudodistance" step to pseudodistances of the form d_φ, that always define pseudodistances, without any loss of generality.

A second consequence is that if $E \subseteq D$ are two sorts, then every definable pseudodistance d on E extends to one on D. Indeed, extend it first in an arbitrary fashion to a formula $\varphi(x, y)$ on $E \times D$. Then d_φ is a pseudodistance on D, and it agrees with d on E.

Remark 1.3. A formula $\psi(x)$ defining the distance to a subset is another property that depends on the structure under consideration, or on its theory. However, we do not know a general construction of definable sets from arbitrary formulas, analogous to that of Remark 1.2, and have good reason to believe that none such exists.

In other words, as far as we know, the set of interpretable sorts depends in a nontrivial way on the theory. This makes it all the more noteworthy that our construction of the universal star sort as D^*_Φ can be carried out in a manner that depends only on the language, and not on the theory.

A *definable map* between two sorts $\sigma : D \to E$ is one whose graph is the zero-set of some formula. Composing a formula with a definable map yields another formula. A special case of such a composition is the formula $d(\sigma(x), y)$, on the product $D \times E$, whose zero-set is indeed the graph of σ. Every formula is uniformly continuous in its arguments, and $d(\sigma(x), y)$ is no exception. It follows that every definable map $\sigma : D \to E$ is uniformly continuous.

Two sorts that admit a definable bijection are, for most intents and purposes (in particular, for those of the present paper) one and the same. Moreover, every sort is in definable bijection with one obtained from the basic sorts by applying each of the operations once, in the given order, so we may pretend that every sort is indeed of this form. Similarly, we may say that a sort D (which may be a basic sort, or one that has already been obtained through some interpretation procedure) is *interpretable* in a family of sorts (E_i) if we can construct from this family (E_i) a sort D' that admits a definable bijection with D.

Consider two languages $\mathcal{L} \subseteq \mathcal{L}'$, where $\mathcal{L}'$ is allowed to add not only symbols, but also sorts. If M' is an $\mathcal{L}'$-structure, and M is the $\mathcal{L}$-structure obtained by dropping the sorts and symbols not present in $\mathcal{L}$, then M is *the $\mathcal{L}$-reduct* of M' and M' is *an $\mathcal{L}'$-expansion* of M. If T' is an $\mathcal{L}'$-theory and T is the collection of $\mathcal{L}$-sentences in T', then T is also the theory of all $\mathcal{L}$-reducts of models of T' (notice, however, that an arbitrary model of T need only admit an elementary extension that is a reduct of a model of T'). In this situation we say that T is *the $\mathcal{L}$-reduct* of T' and that T' is *an $\mathcal{L}'$-expansion* of T.

One special case of an expansion is a *definitional expansion*, in which $\mathcal{L}$ and $\mathcal{L}'$ have the same sorts, and each new symbol of $\mathcal{L}'$ admits an $\mathcal{L}$-definition in T'. In this case, T' is entirely determined by T together with these definitions. A more general case is that of an *interpretational expansion* of T, where T' identifies each new sort of $\mathcal{L}'$ with an interpretable sort of T, and gives $\mathcal{L}$-definitions to all new symbols in $\mathcal{L}'$ (for this to work we also require $\mathcal{L}'$ to contain, in particular, those new symbols that allow T' to identify the new sorts with the corresponding interpretable ones). Again, T, together with the list of interpretations of the new sorts and definitions of the new symbols, determine T'. Moreover, unlike the general situation described in the previous paragraph, here every model of T expands to a model of T'.

Definition 1.4. Let T and T' be two theories, say in disjoint languages. We say that T' is *interpretable* in T if T' is a reduct of an interpretational expansion of T. The two theories are *bi-interpretable* if they admit a common interpretational expansion (which is stronger than just each being interpretable in the other).

A theory has the same sorts (up to a natural identification) as an interpretational expansion. Therefore, somewhat informally, we may say that two theories are bi-interpretable if and only if they have the same sorts.

Let us consider a few more possible constructions of sorts that will become useful at later stages, and show that they can be reduced to the basic construction steps that we allow.

Lemma 1.5. *Let*

$$D_0 \xtwoheadleftarrow{\pi_0} D_1 \xtwoheadleftarrow{\pi_1} \cdots$$

be an inverse system of sorts with surjective definable maps $\pi_n : D_{n+1} \twoheadrightarrow D_n$. Then the inverse limit $D = \varprojlim D_n \subseteq \prod D_n$ is again a sort, which we may equip with the distance

$$d(x, y) = \sum_n (2^{-n} \wedge d(x_n, y_n)) \tag{1}$$

(*or with the restriction of any other definable distance on $\prod D_n$*).

Proof. Indeed, D is the zero-set in $\prod D_n$ of the formula

$$\varphi(x) = \sum_n \big(2^{-n} \wedge d(x_n, \pi_n(x_{n+1}))\big).$$

Let $\varepsilon > 0$, and choose $N \in \mathbf{N}$ large enough depending on ε, and $\delta > 0$ small enough depending on both. Let $a \in \prod D_n$, and assume that $\varphi(a) < \delta$. Since the maps are surjective, there exists $b \in D$ such that $b_N = a_N$. This determines b_n for all $n \leq N$, and having chosen δ small enough, we have $d(a_n, b_n)$ as small as desired for all $n \leq N$. Having chosen N large enough, this yields $d(a, D) \leq d(a, b) < \varepsilon$.

In other words, we have found a formula $\varphi(x)$ that vanishes on D, such that $\varphi(x) < \delta = \delta(\varepsilon)$ implies $\varphi(x, D) < \varepsilon$. This implies that D is a definable subset (see [Ben Yaacov et al. 2008]). □

Proposition 1.6. *Assume that (D_n) is a sequence of sorts, equipped with isometric definable embeddings $D_n \hookrightarrow D_{n+1}$. For convenience, let us pretend these embeddings are the identity map, so $D_0 \subseteq D_1 \subseteq \cdots \subseteq D_n \subseteq \cdots$ is a chain. Assume moreover that the sequence is Cauchy in the Hausdorff distance. In other words, assume that if n is large enough and $n \leq m$, then*

$$d^H(D_n, D_m) = \sup_{x \in D_m} \inf_{y \in D_n} d(x, y)$$

is as small as desired.

Then the completion $E = \widehat{\bigcup D_k}$ is a sort (with definable isometric embedding $D_n \subseteq E$). If $\varphi(x, y)$ is a formula on $E \times F$, for some sort (or product of sorts) F, and φ_n is its restriction to $D_n \times F$, then (φ_n) is an equicontinuous compatible family (by compatible, we mean that each φ_n is the restriction of φ_{n+1}). Conversely, every such family arises from a unique formula on $E \times F$.

Proof. Assume first that we have a large ambient sort E_1 and compatible isometric embeddings $D_n \subseteq E_1$. Since each D_n is a sort, the distance $d(x, D_n) = \inf_{y \in D_n} d(x, y)$ is definable in E_1. By hypothesis, these formulas converge uniformly, and their limit is $d(x, E)$. Then E is a definable subset of E_1, and therefore a sort.

In the general case, we construct E_1 as a quotient of $E_0 = \prod D_n$, whose members we may view as sequences in E. We may freely pass to a subsequence, and assume that $d^H(D_n, D_{n+1}) < 2^{-n-1}$. Say that $a \in E_0$ *converges quickly* if $d(a_n, a_m) \leq 2^{-n} + 2^{-m}$, or equivalently, if $d(a_n, b) \leq 2^{-n}$ where $a_n \to b$ in E. By our hypothesis regarding the rate of convergence of (D_n), every $b \in E$ is the limit of a quickly converging sequence.

Recall the *forced limit* construction from [Ben Yaacov and Usvyatsov 2010]. Formally, it consists of a continuous function $\lim^{\mathrm{F}} : \mathbf{R}^{\mathbf{N}} \to \mathbf{R}$ which is monotone, 1-Lipschitz in the supremum norm on $\mathbf{R}^{\mathbf{N}}$, and most importantly, if $t_n \to s$ fast enough (say $|t_n - s| \leq 2^{-n}$), then $\lim^{\mathrm{F}}(t_n : n \in \mathbf{N}) = s$. We render the expression $\lim^{\mathrm{F}}(t_n : n \in \mathbf{N})$ as $\lim^{\mathrm{F}}{}_{n \to \infty} t_n$, considering it a limit construct. Since $\lim^{\mathrm{F}}$ is continuous, we may apply it to formulas.

Let us fix n, and define on $D_n \times E_0$ a formula

$$\rho_n(x, y) = \lim_{m \to \infty}^{\mathrm{F}} d(x, y_m).$$

If $b \in E_0$ converges quickly to $c \in E$, then $\rho_n(a, b) = d(a, c)$ for every $a \in D_n$. When $b \in E_0$ does not converge quickly (or possibly, at all), the value $\rho_n(a, b)$ is well defined, but potentially meaningless. If $n \leq k$, then ρ_n is the restriction of ρ_k, so we may just denote all of them by ρ.

As in Remark 1.2, we define pseudodistances on E_0 by

$$d_{\rho_n}(y, y') = \sup_{x \in D_n} |\rho(x, y) - \rho(x, y')|.$$

The sequence of formulas (d_{ρ_n}) is increasing. Moreover, if $x, y \in D_n$ and $z \in E_0$, then

$$|\rho(x, z) - \rho(y, z)| \leq \sup_m |d(x, z_m) - d(y, z_m)| \leq d(x, y),$$

so $d_{\rho_n} \leq d_{\rho_{n+1}} \leq d_{\rho_n} + 2^{-n}$. Therefore, the sequence (d_{ρ_n}) converges uniformly to a formula d_ρ on $E_0 \times E_0$, which must define a pseudodistance as well. Let $E_1 = (E_0, d_\rho)$ be the quotient sort. By definition, each $\rho_n(x, y)$ is 1-Lipschitz in y with respect to d_ρ, so it may be viewed as a formula on $D_n \times E_1$. It is also 1-Lipschitz in x with respect to the distance on D_n.

Consider $a \in D_k$ and $b, c \in E_1$, and assume that $b_n \to a$ quickly (but c may be quite arbitrary). We have already observed that $\rho(x, b) = d(x, a)$ for every $x \in D_n$, for every n. Then, for every $n \geq k$,

$$d_{\rho_n}(b, c) = \sup_{x \in D_n} |\rho(x, b) - \rho(x, c)| = \sup_{x \in D_n} |d(x, a) - \rho(x, c)| = \rho(a, c),$$

so $d_\rho(b, c) = \rho(a, c) = \rho_k(a, c)$. If follows that the class of b in E_1 only depends on a. Moreover, the map $\sigma_k : D_k \to E_1$, which sends a to the class of any $b \in E_0$ that converges quickly to a, is definable, by $d_\rho(\sigma_k(x), y) = \rho_k(x, y)$.

If $b, b' \in E_0$ both converge quickly to $a, a' \in D_k$, respectively, then the same reasoning as above yields $d_{\rho_n}(b, b') = d(a, a')$ for every $n \geq k$, and therefore $d_\rho(b, b') = d(a, a')$. Therefore, $\sigma_k : D_k \to E_1$ is an isometric embedding for each k. Since the ρ_k are restrictions of one another, these embeddings are compatible, and we have successfully reduced to the special case treated in the beginning of the proof.

Regarding formulas, the only thing we need to prove is that any compatible equicontinuous family of formulas $\varphi_n(x, y)$ on $D_n \times F$ is the restriction of a formula on $E \times F$. Notice that our hypotheses imply that the formulas φ_n are uniformly bounded, say $|\varphi_n| \leq M$. We may now construct an inverse modulus of continuity, namely a continuous function $\Delta^{-1} : (0, \infty) \to (0, \infty)$ such that $|\varphi_n(x, y) - \varphi_n(x', y)| \leq \Delta^{-1} \circ d(x, x')$ (see [Ben Yaacov and Usvyatsov 2010]; since the family is equicontinuous, we can do this simultaneously for all φ_n). Define on $E \times F$ formulas

$$\psi_n(x, y) = \inf_{x' \in D_n} (\varphi_n(x', y) + \Delta^{-1} \circ d(x, x')).$$

Then ψ_n agrees with φ_n on $D_n \times F$, and equicontinuity together with the convergence of (D_n) in d^H implies that (ψ_n) converges uniformly to a formula $\psi(x, y)$ on $E \times F$ that must extend each φ_n, as claimed. □

It was pointed out by James Hanson that our Proposition 1.6 already appeared in his Ph.D. thesis [Hanson 2020, Proposition 3.4.8]. Similarly, in [Hanson 2020, Remark 3.5.7] he asserts (without proof) something that, to the extent that we understand it (terminology and notation being somewhat nonstandard), is related to our Proposition 2.1.

2. Coding sorts in other sorts

If a and b are two elements in sorts E and D in some structure (model of T), then a is *definable* from b, or lies in the *definable closure* of b — in symbols $a \in \mathrm{dcl}(b)$ — if a is the unique realisation of $\mathrm{tp}(a/b)$ in that structure, as well as in any elementary extension. This implies, and indeed is equivalent to, the predicate $d(x, a)$ being definable with b as parameter, say by a formula $\varphi(x, b)$ (see [Ben Yaacov 2010]).

Let us consider two sorts D and E. In what sense(s) can E be coded in D? A fairly *uniform* fashion for this to happen is if E is interpretable in D, i.e., if it embeds definably in a quotient of $D^{\mathbf{N}}$, or, at the very worst, $D^{\mathbf{N}} \times 2^{\mathbf{N}}$. This would imply a *nonuniform* version: for every $a \in E$ there exists $b \in D^{\mathbf{N}}$ such that $a \in \mathrm{dcl}(b)$. In fact, the converse implication holds as well; this follows fairly easily from Proposition 2.1 below, together with the presentation of $\widehat{\bigcup D_n}$ as a subset of a quotient of $\prod D_n$.

In any case, we want to explore a stronger condition of "nonuniform coding", by singletons in D.

Proposition 2.1. *Let E and D be sorts of a theory T. Assume that for every $a \in E$ (in a model of T) there exists $b \in D$ (possibly in an elementary extension) such that $a \in \mathrm{dcl}(b)$. Then E can be embedded in a limit sort of the form $\widehat{\bigcup D_n}$, as per Proposition 1.6, where each D_n is a quotient of $D \times 2^{\mathbf{N}}$.*

Proof. Consider a type $p \in \mathrm{S}_E(T)$, so $p = \mathrm{tp}(a)$ for some $a \in E$ in a model of T. We may assume that $b \in D$ in the same model is such that $a \in \mathrm{dcl}(b)$, as witnessed by $d(x, a) = \varphi_p(x, b)$.

For $\varepsilon > 0$, let

$$\psi_p(x, y) = \sup_{x'} |d(x, x') - \varphi_p(x', y)|,$$
$$\chi_{p,\varepsilon}(y) = 1 \dot{-} \big(\inf_x \psi_p(x, y)/\varepsilon \dot{-} 1\big).$$

The formula $\psi_p(x, y)$ measures the extent to which $\varphi_p(x', y)$ fails to give us the distance to x. The formula $\chi_{p,\varepsilon}(y)$ tells us whether $x' \mapsto \varphi_p(x', y)$ is close to being the distance to *some* $x \in E$: $\chi_{p,\varepsilon}(y) = 1$ if y codes some x quite well (error less than ε), vanishes if y does not code anything well enough (error at least 2ε), and in all cases its value lies in $[0, 1]$. Of course, $\psi_p(a, b) = 0$, so $\inf_y \psi_p(x, y) < \varepsilon$ defines an open neighbourhood of p.

Let us fix $\varepsilon > 0$ and let p vary. Then the conditions $\inf_y \psi_p(x, y) < \varepsilon$ define an open covering of $S_E(T)$. By compactness, there exists a family $(p_i : i < n)$ such that for every $q \in S_E(T)$, $\inf_y \psi_{p_i}(q, y) < \varepsilon$ for at least one $i < n$. Repeating this, with smaller and smaller ε, we may construct a sequence of types (p_n), as well as $\varepsilon_n \to 0$ such that for every n_0, the open conditions $\inf_y \psi_{p_n}(x, y) < \varepsilon_n$ for $n \geq n_0$ cover $S_E(T)$.

Let $n \in \mathbf{N}$. We view $n = \{0, \dots, n-1\}$ as a quotient of $2^{\mathbf{N}}$, and similarly for $[0, 1]$. Therefore, $D \times n \times [0, 1]$ is a quotient of $D \times 2^{\mathbf{N}}$. For $(x, y, k, t) \in E \times D \times n \times [0, 1]$, define

$$\rho_n(x, y, k, t) = t \cdot \chi_{p_k, \varepsilon_k}(y) \cdot \varphi_{p_k}(x, y).$$

This is indeed a formula, giving rise to a pseudodistance on $D \times n \times [0, 1]$:

$$d_{\rho_n}(y, k, t, y', k', t') = \sup_{x \in E} |\rho_n(x, y, k, t) - \rho_n(x, y', k', t')|.$$

In fact, we may drop n and just write ρ and d_ρ: the only role played by n is being greater than k.

Let D_n be the quotient $(D \times n \times [0, 1], d_\rho)$ (which is, in turn, a quotient of $D \times 2^{\mathbf{N}}$). The inclusion $D \times n \times [0, 1] \subseteq D \times (n+1) \times [0, 1]$ induces an isometric embedding $D_n \hookrightarrow D_{n+1}$. Therefore, in order to show that the hypotheses of Proposition 1.6 are satisfied, all we need to show is that for $n \leq m$ large enough, every member of D_m is close to some member of D_n.

Let $\varepsilon > 0$ be given. Find n_0 such that $\varepsilon_n < \varepsilon$ for $n \geq n_0$. Then, by compactness, find $n_1 > n_0$ such that $\inf_y \psi_{p_n}(x, y) < \varepsilon_n$ for $n_0 \leq n < n_1$ cover $S_E(T)$. Assume now that $n_1 \leq m$, and let $[b, k, t]$ be some class in D_m. If $k < n_1$, then $[b, k, t] \in D_{n_1}$. If $\inf_x \psi_{p_k}(x, b) \geq 2\varepsilon_k$, then $\rho_n(x, b, k, t) = 0$ regardless of x, so $[b, k, t] = [b, 0, 0] \in D_{n_1}$. We may therefore assume that $n_1 \leq k < m$ and there exists $a \in E$ such that $\psi_{p_k}(a, b) < 2\varepsilon_k$.

By our hypothesis regarding the covering of $S_E(T)$, there exists $n_0 \leq \ell < n_1$ such that $\inf_y \psi_{p_\ell}(a, y) < \varepsilon_\ell$. Let $c \in D$ be such that $\psi_{p_\ell}(a, c) < \varepsilon_\ell$, and let $s = t \cdot \chi_{p_k, \varepsilon_k}(b)$. Then

$$\inf_x \psi_{p_\ell}(x, c) < \varepsilon_\ell, \qquad \chi_{p_\ell, \varepsilon_\ell}(c) = 1, \qquad \rho(x, c, \ell, s) = s \cdot \varphi_{p_\ell}(x, c),$$

so

$$\begin{aligned} d_\rho(b, k, t, c, \ell, s) &= s \cdot \sup_x |\varphi_{p_k}(x, b) - \varphi_{p_\ell}(x, c)| \\ &\leq \sup_x |\varphi_{p_k}(x, b) - d(x, a)| + \sup_x |d(x, a) - \varphi_{p_\ell}(x, c)| \\ &= \psi_{p_k}(a, b) + \psi_{p_\ell}(a, c) < 2\varepsilon_k + \varepsilon_\ell < 3\varepsilon. \end{aligned}$$

Then $[c, \ell, s] \in D_{n_1}$ is close enough to $[b, k, t]$. By Proposition 1.6, a limit sort $F = \widehat{\bigcup D_n}$ exists.

Now let us embed $E \hookrightarrow F$. We have already constructed a family (ρ_n) of formulas on $E \times D_n$; let us write them as $\rho_n(x, z)$. Each is 1-Lipschitz in z by definition of the distance on D_n, and they are compatible, so they extend to a formula $\rho(x, z)$ on $E \times F$.

Consider $a \in E$, and let $\varepsilon > 0$. As above, there exists ℓ such that $\varepsilon_\ell < \varepsilon$, and $c \in D$ such that $\psi_{p_\ell}(a, c) < \varepsilon_\ell$. Let $a' = [c, \ell, 1] \in D_{\ell+1} \subseteq F$. Again, as above, $\chi_{p_\ell,\varepsilon_\ell}(c) = 1$, so $\rho(x, a') = \varphi_{p_\ell}(x, c)$, and

$$\sup_x |d(x, a) - \rho(x, a')| = \sup_x |d(x, a) - \varphi_{p_\ell}(x, c)| = \psi_{p_\ell}(a, c) < \varepsilon_\ell < \varepsilon.$$

Doing this with $\varepsilon \to 0$, we obtain a sequence (a_n) in F such that $\rho(x, a_n)$ converges uniformly to $d(x, a)$. By definition of the distance on F as d_ρ, this sequence is Cauchy, with limit $\tilde{a} \in F$, say, and $\rho(x, \tilde{a}) = d(x, a)$. In particular, for $z \in F$,

$$d(z, \tilde{a}) = \sup_x |\rho(x, z) - \rho(x, \tilde{a})| = \sup_x |\rho(x, z) - d(x, a)|,$$

so $a \mapsto \tilde{a}$ is definable. By the same reasoning, if $a, a' \in E$, then

$$d(\tilde{a}, \tilde{a}') = \sup_x |\rho(x, \tilde{a}) - \rho(x, \tilde{a}')| = \sup_x |d(x, a) - d(x, a')| = d(a, a'),$$

so the embedding is isometric, completing the proof. □

Remark 2.2. A closer inspection of the proof can yield a necessary and sufficient condition (but we shall not use this): A sort E can be embedded in a limit sort of the form $\widehat{\bigcup D_n}$, where each D_n is a quotient of $D \times 2^{\mathbf{N}}$, if and only if, for every $a \in E$ and $\varepsilon > 0$, there exists $b \in D$ and a formula $\varphi(x, b)$ that approximates $d(x, a)$ with error at most ε.

In Proposition 2.1, we cannot replace $D \times 2^{\mathbf{N}}$ with just D (if D is a singleton, then any increasing union of quotients of D is a singleton, and yet $E = \{0, 1\}$ satisfies the hypothesis of Proposition 2.1). Instead, let us prove that this does not change much, in the sense that formulas on $D \times 2^{\mathbf{N}}$ or on just D are almost the same thing.

Lemma 2.3. *Let D and E be sorts, and let $\varphi(x, t, y)$ be a formula on $D \times 2^{\mathbf{N}} \times E$. Then φ can be expressed as a uniform limit of continuous combinations of formulas on $D \times E$ and on $2^{\mathbf{N}}$ separately (recalling that formulas on $2^{\mathbf{N}}$ are just continuous functions $2^{\mathbf{N}} \to \mathbf{R}$).*

Proof. For $n \in \mathbf{N}$ and $k \in 2^n$, let $\delta_{n,k}(t) = 1$ if t extends k, and 0 otherwise. Let also $\tilde{k} \in 2^{\mathbf{N}}$ be the extension of k by zeros, and $\varphi_{n,k}(x, y) = \varphi(x, \tilde{k}, y)$.

Then $\varphi_{n,k}$ is a formula on $D \times E$ and $\delta_{n,k}$ is a formula on $2^{\mathbf{N}}$, so we may define a formula

$$\varphi_n(x, t, y) = \sum_{k \in \{0,1\}^n} \delta_{n,k}(t)\varphi_{n,k}(x, y).$$

Since $\varphi(x, t, y)$ is uniformly continuous in t, $\varphi_n \to \varphi$ uniformly. □

Definition 2.4. Let T be a theory, D a sort, and $D^0 \subseteq D$ a definable subset (or even type-definable, namely, the zero-set of a formula). We say that D is a *coding sort*, with *exceptional set* D^0, if the following hold:

(i) *Coding models*: if $M \models T$ and $a \in D(M) \smallsetminus D^0(M)$, then there exists $N \preceq M$ such that $\operatorname{dcl}(a) = \operatorname{dcl}(N)$. We then say that *$a$ codes N*.

(ii) *Density*: if $M \models T$ is separable, then the set of those $a \in D(M) \smallsetminus D^0(M)$ that code M is dense in $D(M)$.

We may denote a coding sort by D alone, considering D^0 as implicitly given together with D.

The need for an exceptional set will arise at a later stage; for the time being, we are simply going to ensure that its presence does not cause any trouble.

Definition 2.5. Let T be a theory in a language $\mathcal{L}$, and let D be a coding sort for T.

We define a single-sorted language $\mathcal{L}_{2D}$ to consist of a binary predicate symbol for each formula on $D \times D$ (possibly restricting this to a dense family of such formulas). We define T_{2D} as the $\mathcal{L}_{2D}$-theory of D — namely, the theory of all $D(M)$, viewed naturally as $\mathcal{L}_{2D}$-structures, where M varies over models of T.

Clearly, T_{2D} is interpretable from T. The 2 is there to remind us that only binary predicates on D are named in the language.

Our aim, in the end, is to recover from a groupoid the theory of some coding sort D, and show that it is bi-interpretable with T. In particular, we need to recover the definable predicates on D from the groupoid. In [Ben Yaacov 2022] we managed to recover predicates of all arities, at the price of some additional work. In the present paper we choose to follow a different path, recovering only binary predicates (i.e., only T_{2D}), and instead show that these suffice.

Proposition 2.6. *Let T be a theory in a language $\mathcal{L}$, and let D be a coding sort for T. Then T_{2D} is bi-interpretable with T.*

Proof. Consider T', obtained from T by adjoining D as a new sort, and naming the full induced structure. It is, by definition, an interpretational expansion of T, and it suffices to show that it is also an interpretational expansion of T_{2D}.

By Lemma 2.3, every formula on $(D \times 2^{\mathbf{N}}) \times (D \times 2^{\mathbf{N}})$ is definable in T_{2D}. In particular, every quotient of $D \times 2^{\mathbf{N}}$ is interpretable in T_{2D}, as is every embedding of one such quotient in another. Therefore, if (D_n) is an increasing chain of quotients of $D \times 2^{\mathbf{N}}$ that converges in the sense of Proposition 1.6, then $E = \widehat{\bigcup D_n}$ is interpretable in T_{2D}.

Consider now a sort E of T. Every member of E belongs to a separable model of T and is therefore definable from a member of D. By Proposition 2.1, we may embed E in a sort $\widetilde{E}$ which is of the form $\widehat{\bigcup D_n}$, for appropriate quotients of $D \times 2^{\mathbf{N}}$, as in the previous paragraph. This presentation of E need not be unique, so let us just fix one such.

Say E' is another sort of T, so $E' \subseteq \widetilde{E}' = \widehat{\bigcup D'_n}$ as above. Any formula on $\widetilde{E} \times \widetilde{E}'$ is, by Proposition 1.6, coded by a sequence of formulas on $D_n \times D'_n$ (its restrictions), i.e., by formulas on $(D \times 2^{\mathbf{N}})^2$. It is therefore definable in T_{2D}. In particular, the distance to (the copy of) E in $\widetilde{E}$ is definable in T_{2D}, so each sort E of T can be interpreted in T_{2D} (or at least, some isometric copy of E is interpretable). Similarly, every formula on $E \times E'$, can be extended to a formula on $\widetilde{E} \times \widetilde{E}'$, so it is definable in T_{2D} (on the copies of E and E').

Consider now a finite product $E = \prod_{i<n} E_i$ of sorts of T. We have already chosen embeddings $E \subseteq \widetilde{E}$ and $E_i \subseteq \widetilde{E}_i$ as above. The projection map $\pi_i : E \to E_i$ can be coded by a formula on $E \times E_i$, namely

$$\Gamma_{\pi_i}(x, y) = d_{E_i}(x_i, y),$$

where Γ stands for "graph". We have already observed that such a formula is definable in T_{2D}. It follows that the structure of E as a product of the E_i is definable in T_{2D}. Finally, any formula on $E_0 \times \cdots \times E_{n-1}$ can be viewed as a unary formula on the product E, which is, again, definable in T_{2D}.

In conclusion, we can interpret every sort of T in T_{2D}, and recover the full structure on these sorts. In other words, T' is indeed an interpretational expansion of T_{2D}, completing the proof. □

3. Groupoid constructions and reconstruction strategies

In this section we propose a general framework for "reconstruction theorems". To any coding sort D (see Definition 2.4) we associate a topological groupoid $\boldsymbol{G}_D(T)$ from which the theory T_{2D} of Proposition 2.6 can be reconstructed. Since T is bi-interpretable with T_{2D}, the groupoid $\boldsymbol{G}_D(T)$ determines the bi-interpretation class of T. If the coding sort is moreover determined by the bi-interpretation class of T (up to definable bijection), then the groupoid is a bi-interpretation invariant. Various previously known constructions fit in this framework, as well as the one towards which the present paper aims.

For a general treatment of topological groupoids, we refer the reader to [Mackenzie 1987], or, for the bare essentials we need here, to [Ben Yaacov 2022]. We recall that a *groupoid* $\boldsymbol{G}$ is defined either as a small category in which all morphisms are invertible, or algebraically, as a single set (of all morphisms), equipped with a partial composition law and a total inversion map, satisfying appropriate axioms. When viewed as a category, the set of objects can be identified with the set of identity morphisms, and we call it the *basis* $\boldsymbol{B}$ of $\boldsymbol{G}$. In the algebraic formalism, which we follow here, the basis is $\boldsymbol{B} = \{e \in \boldsymbol{G} : e^2 = e\} \subseteq \boldsymbol{G}$. If $g \in \boldsymbol{G}$, then $s(g) = g^{-1}g$ and

$t(g) = gg^{-1}$ are both defined, and belong to $\boldsymbol{B}$, being the *source* and *target* of g, respectively. The domain of the composition law is

$$\mathrm{dom}(\,\cdot\,) = \{(g, h) : s(g) = t(h)\} \subseteq \boldsymbol{G}^2.$$

A *topological groupoid* is a groupoid equipped with a topology in which the partial composition law and total inversion map are continuous. In a topological groupoid the source and target maps $s, t : \boldsymbol{G} \to \boldsymbol{B}$ are continuous as well, $\boldsymbol{B}$ is closed in $\boldsymbol{G}$, and $\mathrm{dom}(\,\cdot\,)$ closed in $\boldsymbol{G}^2$. A topological groupoid $\boldsymbol{G}$ is *open* if, in addition, the composition law $\cdot : \mathrm{dom}(\cdot) \to \boldsymbol{G}$ is open, or equivalently, if the source map $s : \boldsymbol{G} \to \boldsymbol{B}$ (or target map $t : \boldsymbol{G} \to \boldsymbol{B}$) is open.

A (topological) group is a (topological) groupoid whose basis is a singleton. Such a topological groupoid is always open.

Definition 3.1. Let T be a theory in a countable language, and D a coding sort. We let $\mathrm{S}_{D\times D}(T)$ denote the space of types of pairs of elements of D. We define the following two subsets of $\mathrm{S}_{D\times D}(T)$:

$$\boldsymbol{G}^0_D(T) = \{\mathrm{tp}(a, a) : a \in D^0\},$$
$$\boldsymbol{G}_D(T) = \boldsymbol{G}^0_D(T) \cup \{\mathrm{tp}(a, b) : a, b \in D \smallsetminus D^0 \text{ and } \mathrm{dcl}(a) = \mathrm{dcl}(b)\},$$

where a and b vary over all members of D (or D^0) in models of T. We equip $\boldsymbol{G}_D(T)$ with the induced topology, as well as with the following inversion law and partial composition law:

$$\mathrm{tp}(a, b)^{-1} = \mathrm{tp}(b, a), \qquad \mathrm{tp}(a, b) \cdot \mathrm{tp}(b, c) = \mathrm{tp}(a, c).$$

We also write $\boldsymbol{B}_D(T)$ for $\mathrm{S}_D(T)$, and identify $\mathrm{tp}(a) \in \boldsymbol{B}_D(T)$ with $\mathrm{tp}(a, a) \in \boldsymbol{G}_D(T)$. This identifies $\boldsymbol{B}^0_D(T) = \mathrm{S}_{D^0}(T)$ with $\boldsymbol{G}^0_D(T)$.

Notice that the density hypothesis in Definition 2.4 implies that $\boldsymbol{G}_D(T)$ is dense in $\mathrm{S}_{D\times D}(T)$.

Convention 3.2. We usually consider the theory T and the coding sort D to be fixed and drop them from notation, so $\boldsymbol{G} = \boldsymbol{G}_D(T)$, $\boldsymbol{B} = \boldsymbol{B}_D(T)$, and so on.

Lemma 3.3. *Let D be a coding sort for T.*

(i) *As defined above $\boldsymbol{G} = \boldsymbol{G}_D(T)$ is a Polish open topological groupoid with basis $\boldsymbol{B} = \boldsymbol{B}_D(T)$.*

(ii) *If $g = \mathrm{tp}(a, b) \in \boldsymbol{G}$, then $s(g) = \mathrm{tp}(b) \in \boldsymbol{B}$ is its source, and $t(g) = \mathrm{tp}(a) \in \boldsymbol{B}$ its target.*

(iii) *If d is a definable distance on D, then the family of sets*

$$U_r = \{\mathrm{tp}(a, b) \in \boldsymbol{G} : d(a, b) < r\},$$

for $r > 0$, forms a basis of open neighbourhoods for $\boldsymbol{B}$ in $\boldsymbol{G}$.

Proof. It is easy to check that $\boldsymbol{G}$ is a topological groupoid with basis $\boldsymbol{B}$ and the stated source and target. Since the language is countable, the space $\mathrm{S}_{D\times D}(T)$ is compact metrisable, and therefore Polish. As a condition on $\mathrm{tp}(a, b)$, the property $\mathrm{dcl}(a) = \mathrm{dcl}(b)$ is G_δ by [Ben Yaacov 2022, Lemma 5.1], and $a, b \notin D^0$ is open. Therefore $\boldsymbol{G}$ is Polish, as the union of a closed subset and a G_δ subset of a Polish space.

Each set U_r is open and contains $\boldsymbol{B}$. On the other hand, if U is any open neighbourhood of $\boldsymbol{B}$ in $\boldsymbol{G}$, then it must be of the form $W \cap \boldsymbol{G}$, where W is an open neighbourhood of $\boldsymbol{B}$ in $\mathrm{S}_{D\times D}(T)$. Since $\boldsymbol{B}$ is defined there by the condition $d(x, y) = 0$, and by compactness, W must contain $[d(x, y) < r]$ for some $r > 0$, so U contains U_r.

It is left to show that the target map $t : \boldsymbol{G} \to \boldsymbol{B}$ is open. First, consider $g \in \boldsymbol{G} \smallsetminus \boldsymbol{G}^0 \subseteq \mathrm{S}_{D\times D}(T)$. Let $[x \in D^0] \subseteq \mathrm{S}_{D\times D}(T)$ be the set of types $p(x, y)$ that imply $x \in D^0$, and similarly for y, observing that $g \notin [x \in D^0] \cup [y \in D^0]$. Since this union is a closed set, g admits a basis of neighbourhoods in $\mathrm{S}_{D\times D}(T)$ that are disjoint from $[x \in D^0] \cup [y \in D^0]$. By Urysohn's lemma and the identification of formulas with continuous functions on types, g admits a basis of neighbourhoods of the form $[\varphi(x, y) > 0]$, where $\varphi(x, y)$ vanishes if $x \in D^0$ or $y \in D^0$. The family of sets $[\varphi(x, y) > 0] \cap \boldsymbol{G}$ for such φ is a basis of neighbourhoods for g in $\boldsymbol{G}$.

Assume we are given such a neighbourhood $g \in U = [\varphi(x, y) > 0] \cap \boldsymbol{G}$ (so $\varphi(x, y)$ vanishes if $x \in D^0$ or $y \in D^0$). Let $V = [\sup_y \varphi(x, y) > 0] \subseteq \mathrm{S}_D(T) = \boldsymbol{B}$. Then V is open, and clearly $t(U) \subseteq V$. Conversely, assume that $\mathrm{tp}(a) \in V$, where $a \in D(M)$ for some $M \vDash T$. Then there exists $b \in D(M)$ such that $\varphi(a, b) > 0$. By hypothesis on φ, it follows that $a, b \notin D^0$. In particular, a codes a separable $N \preceq M$, and we may assume that $b \in D(N)$. Now, by the density property and the uniform continuity of φ, we may assume that b also codes N, so $\mathrm{tp}(a, b) \in U$. This proves that $t(U) = V$.

Now let $g = \mathrm{tp}(a, a) \in \boldsymbol{G}^0$. We have a basis of neighbourhoods of g in $\boldsymbol{G}$ consisting of sets of the form

$$U = [\varphi(x) > 0] \cap [d(x, y) < r] \cap \boldsymbol{G},$$

where $\varphi(a) > 0$. It is then easily checked that $t(U) = [\varphi(x) > 0]$, since we may always take $y = x$ as witness. This completes the proof. □

Definition 3.4. Let $\boldsymbol{G}$ be a topological groupoid. Say that a function $\varphi : \boldsymbol{G} \to \mathbf{R}$ is *uniformly continuous and continuous* (*UCC*) if it is continuous on $\boldsymbol{G}$, and in addition satisfies the following uniform continuity condition: for every $\varepsilon > 0$ there exists an open neighbourhood U of the basis $\boldsymbol{B}$ such that for every $g \in \boldsymbol{G}$,

$$h \in UgU \quad\Longrightarrow\quad |\varphi(g) - \varphi(h)| < \varepsilon.$$

Notice that unlike the situation for groups, the uniform continuity condition does not imply continuity (it is quite possible that $g_n \to h$ while $h \notin \boldsymbol{G} g_n \boldsymbol{G}$ for any n).

Proposition 3.5. *Assume that D is a coding sort for T, and let $\boldsymbol{G} = \boldsymbol{G}_D(T)$. Let $\varphi(x, y)$ be a formula on $D \times D$, and let $\varphi_{\boldsymbol{G}} : \boldsymbol{G} \to \mathbf{R}$ be the naturally induced function*

$$g = \operatorname{tp}(a, b) \quad \Longrightarrow \quad \varphi_{\boldsymbol{G}}(g) = \varphi(a, b).$$

Then the map $\varphi \mapsto \varphi_{\boldsymbol{G}}$ defines a bijection between formulas on $D \times D$, up to equivalence, and UCC functions on $\boldsymbol{G}$.

Proof. Let us first check that if φ is a formula, then $\varphi_{\boldsymbol{G}}$ is UCC. It is clearly continuous. The uniform continuity condition follows from the fact that φ is uniformly continuous in each argument, together with the fact that for any $\delta > 0$ we may take choose $U = [d(x, y) < \delta] \cap \boldsymbol{G}$.

Conversely, assume that $\psi : \boldsymbol{G} \to \mathbf{R}$ is UCC. By density, the function ψ admits at most one continuous extension to $\mathrm{S}_{D\times D}(T)$, and we need to show that one such exists. In other words, given $p \in \mathrm{S}_{D\times D}(T)$ and $\varepsilon > 0$, it suffices to find a neighbourhood $p \in V \subseteq \mathrm{S}_{D\times D}(T)$ such that ψ varies by less than ε on $V \cap \boldsymbol{G}$. This is easy if $p \in \boldsymbol{G}$, so we may assume that $p \notin \boldsymbol{G}$.

Let us fix $\varepsilon > 0$ first. By uniform continuity of ψ and Lemma 3.3(iii), there exists $\delta > 0$ such that $|\psi(g) - \psi(ugv)| < \varepsilon$ whenever $g \in \boldsymbol{G}$, $u, v \in [d(x, y) < \delta] \cap \boldsymbol{G}$, and ugv is defined.

Given $p = \operatorname{tp}(a_0, b_0)$, we may assume that $a_0, b_0 \in D(M)$ for some separable model M. Since $p \notin \boldsymbol{G}$, we must have $a_0 \neq b_0$, and (possibly decreasing δ) we may assume that $d(a_0, b_0) > 2\delta$. By the density property, there exist $a_1, b_1 \in D(M)$ that code M, with $d(a_0, a_1) + d(b_0, b_1) < \delta$, so $d(a_1, b_1) > \delta$. Let $g_1 = \operatorname{tp}(a_1, b_1) \in \boldsymbol{G}$. By continuity, there exists an open neighbourhood $g_1 \in V_1 \subseteq \mathrm{S}_{D\times D}(T)$ such that $|\psi(g_1) - \psi(h)| < \varepsilon$ for every $h \in V_1 \cap \boldsymbol{G}$. Possibly decreasing V_1, we may further assume that $\operatorname{tp}(a, b) \in V_1$ implies $d(a, b) > \delta$. We may even assume that V_1 is of the form $[\chi < \delta]$, where $\chi(x, y) \geq 0$ is a formula and $\chi(a_1, b_1) = \chi(g_1) = 0$. Define

$$\chi'(x, y) = \inf_{x', y'} [d(x, x') + d(y, y') + \psi(x', y')],$$

$$V = [\chi'(x, y) < \delta] \subseteq \mathrm{S}_{D\times D}(T).$$

Then V is open, $p \in V$, and $\operatorname{tp}(a, b) \in V$ implies $a \neq b$ (in other words, $V \cap \boldsymbol{B} = \varnothing$).

In order to conclude, consider any $g_2 = \operatorname{tp}(a_2, b_2) \in V \cap \boldsymbol{G}$. Since $a_2 \neq b_2$, they cannot belong to the exceptional set, so both code some separable model N. By definition of V, there exist $a_3, b_3 \in D(N)$ such that

$$\chi(a_3, b_3) + d(a_2, a_3) + d(b_2, b_3) < \delta.$$

By the density property, and uniform continuity of χ, we may assume that a_3 and b_3 code N as well. Let $g_3 = \mathrm{tp}(a_3, b_3)$, $u = \mathrm{tp}(a_3, a_2)$, $v = \mathrm{tp}(b_2, b_3)$. Then $g_3 = ug_2v \in V_1$, so

$$|\psi(g_2) - \psi(g_1)| \leq |\psi(g_2) - \psi(g_3)| + |\psi(g_3) - \psi(g_1)| \leq 2\varepsilon.$$

Therefore, ψ varies by less than 4ε on $V \cap \boldsymbol{G}$, which is good enough. □

Corollary 3.6. *Every UCC function on $\boldsymbol{G}_D(T)$ is bounded.*

Definition 3.7. Let $\boldsymbol{G}$ be a groupoid. A *seminorm* on $\boldsymbol{G}$ is a function $\rho : \boldsymbol{G} \to \mathbf{R}^+$ that satisfies

- $\rho{\restriction}_{\boldsymbol{B}} = 0$,
- $\rho(g) = \rho(g^{-1})$, and
- $\rho(gh) \leq \rho(g) + \rho(h)$, when defined.

It is a *norm* if $\rho(g) = 0$ implies $g \in \boldsymbol{B}$.

A norm ρ is *compatible* with a topology on $\boldsymbol{G}$ if it is continuous, and the sets

$$\{\rho < r\} = \{g \in \boldsymbol{G} : \rho(g) < r\},$$

for $r > 0$, form a basis of neighbourhoods for $\boldsymbol{B}$.

Corollary 3.8. *The correspondence of Proposition 3.5 restricts to a one-to-one correspondence between definable distances d on D and compatible norms on $\boldsymbol{G} = \boldsymbol{G}_D(T)$.*

Proof. Let d be a definable distance on $D \times D$ and ρ_d the corresponding UCC function $\boldsymbol{G}$. Then ρ_d is clearly a continuous norm, and it is a compatible norm by Lemma 3.3(iii).

The converse is more delicate. Let ρ be a compatible norm. Then it is continuous, and it is easy to see that every continuous seminorm is UCC, so $\rho = \varphi_{\boldsymbol{G}}$ (in the notations of Proposition 3.5) for some formula $\varphi(x, y)$. If $a, b, c \in D$ all code the same separable model, then $\varphi(a, a) = 0$ and $\varphi(a, b) \leq \varphi(a, c) + \varphi(b, c)$. The set of types of such triplets is dense in $\mathrm{S}_{D\times D\times D}(T)$, by the density property, so the same holds throughout and φ defines a pseudodistance.

It is left to show that φ defines a distance (and not merely a pseudodistance). Let d be any definable distance on D, say the one distinguished in the language. We already know that ρ_d is a compatible norm. Therefore, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\{\rho < \delta\} \subseteq \{\rho_d < \varepsilon\}$. As in the previous paragraph, this means that the (closed) condition $\varphi(a, b) < \delta \implies d(a, b) \leq \varepsilon$ holds on a dense set of types, and therefore throughout. In particular, if $\varphi(a, b) = 0$, then $a = b$, and the proof is complete. □

Let T be a theory, D a coding sort for T, and $\boldsymbol{G} = \boldsymbol{G}_D(T)$. Then from $\boldsymbol{G}$, given as a topological groupoid, we can essentially recover the language $\mathcal{L}_D$ and the theory T_{2D} follows.

(i) Choose an arbitrary compatible norm ρ on $\boldsymbol{G}$ (which exists, by Corollary 3.8).

(ii) Let $\mathcal{L}_{\boldsymbol{G}}$ consists of a single sort, also named D, together with a binary predicate symbol P_ψ for each UCC function ψ on $\boldsymbol{G}$. We know that ψ is bounded (Corollary 3.6), and we impose the same bound on P_ψ. We also know that for every $\varepsilon > 0$ there exists a neighbourhood U of $\boldsymbol{B}$ such that $h \in UgU$ implies $|\psi(g) - \psi(h)| < \varepsilon$, and since ρ is compatible, there exists $\delta = \delta_\psi(\varepsilon) > 0$ such that the same holds when $U = \{\rho < \delta\}$. We then impose the corresponding modulus of uniform continuity on P_ψ, namely requiring that

$$d(x, x') \vee d(y, y') < \delta_\psi(\varepsilon) \quad \Longrightarrow \quad |P_\psi(x, y) - P_\psi(x', y')| \leq \varepsilon.$$

We also use the bound on ρ as bound on the distance predicate.

(iii) Let us fix $e \in \boldsymbol{B}$, and consider the set

$$e\boldsymbol{G} = \{g \in \boldsymbol{G} : t_g = e\}.$$

If $g, h \in e\boldsymbol{G}$, then $g^{-1}h$ is defined, and for any UCC ψ we let

$$P_\psi(g, h) = \psi(g^{-1}h).$$

In particular, $d(g, h) = P_\rho(g, h) = \rho(g^{-1}h)$ is a distance function on $e\boldsymbol{G}$.

Assume now that $g', h' \in e\boldsymbol{G}$ as well, and $d(g, g') \vee d(h, h') < \delta = \delta_\psi(\varepsilon)$. Let $u = g'^{-1}g$ and $v = h^{-1}h'$. Then $g'^{-1}h' = ug^{-1}hv$, and $u, v \in \{\rho < \delta\}$, so indeed

$$|P_\psi(g, h) - P_\psi(g', h')| \leq \varepsilon,$$

as required. The bounds are also respected, so $e\boldsymbol{G}$, equipped with the distance and interpretations of P_ψ, is an $\mathcal{L}_{\boldsymbol{G}}$-prestructure, and its completion $\widehat{e\boldsymbol{G}}$ is an $\mathcal{L}_{\boldsymbol{G}}$-structure.

(iv) We define $T_{\boldsymbol{G}}$ as the theory of the collection of all $\mathcal{L}_{\boldsymbol{G}}$-structures of this form:

$$T_{\boldsymbol{G}} = \mathrm{Th}_{\mathcal{L}_{\boldsymbol{G}}}\big(\widehat{e\boldsymbol{G}} : e \in \boldsymbol{B}\big).$$

By "essentially recover", we mean the following.

Theorem 3.9. *Let T be a theory, D a coding sort for T, and $\boldsymbol{G} = \boldsymbol{G}_D(T)$. Let $\mathcal{L}_{\boldsymbol{G}}$ and $T_{\boldsymbol{G}}$ be constructed as in the preceding discussion. Then $T_{\boldsymbol{G}}$ and T_{2D} are one and the same, up to renaming the binary predicate symbols, and up to an arbitrary choice of the distance on the sort D (from among all definable distances).*

In particular, this procedure allows us to recover from $\boldsymbol{G}$ a theory $T_{\boldsymbol{G}}$ that is bi-interpretable with T.

Proof. By Corollary 3.8, step (i) consists exactly of choosing a definable distance d on D, and the corresponding norm $\rho = d_{\boldsymbol{G}}$. This choice is irremediably arbitrary. By Proposition 3.5, in step (ii) there is a natural bijection between symbols of $\mathcal{L}_D$ (corresponding to formulas $\varphi(x, y)$ on $D \times D$, up to equivalence) and symbols of $\mathcal{L}_{\boldsymbol{G}}$: to φ we associate the UCC function $\psi_\varphi = \varphi_{\boldsymbol{G}}$, to which in turn we associate the symbol P_{ψ_φ}.

Finally, let $M \vDash T$ be separable, $a \in D(M)$ a code for M, and $e = \operatorname{tp}(a) \in \boldsymbol{B}$. Let $D(M)_1$ denote the set of $b \in D(M)$ that also code M. If $b \in D(M)_1$, then $g_b = \operatorname{tp}(a, b) \in e\boldsymbol{G}$. Moreover, if $b, c \in D(M)_1$ and φ is a formula on $D \times D$, then $\operatorname{tp}(b, c) = g_b^{-1} g_c \in \boldsymbol{G}$, so

$$\varphi(b, c) = \psi_\varphi(g_b^{-1} g_c) = P_{\psi_\varphi}(g_b, g_c).$$

In particular, $d(b, c) = d(g_b, g_c)$ (where the first is the distance we chose on D, and the second the distance we defined on $e\boldsymbol{G}$ in step (iii)). Thus, up to representing φ by the symbol P_{ψ_φ}, the map $b \mapsto g_b$ defines an isomorphism of the $\mathcal{L}_D$-prestructure $D(M)_1$ with the $\mathcal{L}_{\boldsymbol{G}}$-prestructure $e\boldsymbol{G}$. This extends to an isomorphism of the respective completions: $D(M) \simeq \widehat{e\boldsymbol{G}}$.

It follows that, up to this change of language (and choice of distance), the theory $T_{\boldsymbol{G}}$ defined in step (iv) is the theory of all separable models of T_{2D}. Since T is in a countable language, T_{2D} is in a "separable language", so it is equal to the theory of all its separable models.

By Proposition 2.6, T is bi-interpretable with T_{2D}, and therefore also with $T_{\boldsymbol{G}}$. $\square$

Having achieved this, we are ready to start producing reconstruction theorems: all we need is a coding sort that only depends (up to definable bijection) on the bi-interpretation class of T.

Example 3.10. Let T be an $\aleph_0$-categorical theory. Let M be its unique separable model, and let a be any sequence (possibly infinite, but countable), in any sort or sorts, such that $\operatorname{dcl}(a) = \operatorname{dcl}(M)$ (for example, any dense sequence will do). Let $D_{T,0}$ be the set of realisations of $p = \operatorname{tp}(a)$. Since T is $\aleph_0$-categorical, $D_{T,0}$ is a definable set, i.e., a sort. It is easy to check that it is a coding sort (with no exceptional set).

If b is another code for M, and $D'_{T,0}$ is the set of realisations of $\operatorname{tp}(b)$, then $\operatorname{dcl}(a) = \operatorname{dcl}(b)$ and $\operatorname{tp}(a, b)$ defines the graph of a definable bijection $D_{T,0} \simeq D'_{T,0}$. Therefore, $D_{T,0}$ does not depend on the choice of a. Moreover, assume that T' is an interpretational expansion of T. Then it has a model M' that expands M accordingly. But then $\operatorname{dcl}(M') = \operatorname{dcl}(M) = \operatorname{dcl}(a)$ (as calculated when working in T'), so $D_{T',0} = D_{T,0}$. It follows that $D_{T,0}$ only depend on the bi-interpretation class of T.

Since $\mathrm{S}_{D_{T,0}}(T) = \{p\}$ is a singleton, the groupoid

$$G(T) = \boldsymbol{G}_{D_{T,0}}(T)$$

is in fact a group. It only depends on the bi-interpretation class of T (since $D_{T,0}$ only depends on it) and by Theorem 3.9, it is a complete bi-interpretation invariant for T.

We leave it to the reader to check that

$$G(T) \simeq \mathrm{Aut}(M),$$

and that the reconstruction result is just a complicated restatement of those of [Ahlbrandt and Ziegler 1986; Ben Yaacov and Kaïchouh 2016].

Example 3.11. Let T be a theory in classical logic. In [Ben Yaacov 2022], using an arbitrary parameter Φ, we gave an explicit construction of a set of infinite sequences D_Φ. We showed that it is a definable set in the sense of continuous logic, and that its interpretation in models of T only depend on the bi-interpretation class of T (up to a definable bijection). It also follows from what we showed that it is a coding sort (without exceptional set). Since it is unique, let us denote it by D_T (in fact, we could also just denote it by D: its construction only depends on the language, and then we simply restrict our consideration of it to models of T). We then proved that the groupoid

$$\boldsymbol{G}(T) = \boldsymbol{G}_{D_T}(T)$$

is a complete bi-interpretation invariant for T. This is a special case of Theorem 3.9.

Example 3.12. Let T be a (complete) theory in continuous logic. In [Ben Yaacov 2022] we defined when a sort D_T is a *universal Skolem sort*, and proved that if such a sort exists, then it is unique, and only depends on the bi-interpretation class of T (in contrast with the previous example, here we do not have a general construction for such a sort, let alone a uniform one, so it really does depend on T). We proved that if T admits a universal Skolem sort D_T, then

$$\boldsymbol{G}(T) = \boldsymbol{G}_{D_T}(T)$$

is a complete bi-interpretation invariant for T.

Again, we also proved that D_T is a coding sort, so this is a special case of Theorem 3.9.

Remark 3.13. Example 3.12 encompasses the two previous examples in the following sense.

- If T is classical, then the sort D_T of Example 3.11 is a universal Skolem sort, so Example 3.11 is a special case of Example 3.12.
- If T is $\aleph_0$-categorical, then $D_T = D_{T,0} \times 2^{\mathbf{N}}$ is a universal Skolem sort, so

$$\boldsymbol{G}(T) \simeq 2^{\mathbf{N}} \times G(T) \times 2^{\mathbf{N}}, \quad \text{with groupoid law } (\alpha, g, \beta) \cdot (\beta, h, \gamma) = (\alpha, gh, \gamma).$$

Consequently, $\boldsymbol{B}(T) = 2^{\mathbf{N}}$, and if $e \in \boldsymbol{B}(T)$, then $G(T) \simeq e\boldsymbol{G}(T)e$. Therefore, the reconstruction of Example 3.10 can be recovered from a special case of Example 3.12.

In both Example 3.11 and Example 3.12, the basis $\mathrm{S}_{D_T}(T)$ is homeomorphic to the Cantor space $2^{\mathbf{N}}$.

However, in [Ben Yaacov 2022] we also gave an example of a continuous theory which does not admit a universal Skolem sort. In particular, the explicit construction of D_T as D_Φ in the case of a classical theory simply does not extend, as is, to continuous logic. The rest of this article is dedicated to presenting a modified version of this construction, giving rise to a coding sort that *does* have an exceptional set (a very simple one, consisting of a single point), allowing us to prove a reconstruction theorem for every first-order theory in a countable language (in continuous or classical logic).

4. Star spaces

Before we can construct our coding sort, we require a technical detour, where we introduce star sets in general, and, in the model-theoretic context, star sorts. For the time being, we must ask the reader to bear with us — the usefulness of these notions for our goal is explained in some detail at the beginning of Section 6.

Definition 4.1. (i) A *retraction set* is a set X equipped with an action of the multiplicative monoid $[0, 1]$. In particular, $1 \cdot x = x$ for all $x \in X$, and $\alpha(\beta x) = (\alpha\beta)x$ (so this is a little stronger than a homotopy).

(ii) It is a *star set* if $0 \cdot x$ does not depend on x. We then denote this common value by $0 \in X$, and call it the *root* of X.

(iii) A *topological retraction* (*star*) *space* is one equipped with a topology making the action $[0, 1] \times X \to X$ continuous.

(iv) A *metric star space* is one equipped with a distance function satisfying $d(\alpha x, \alpha y) \leq \alpha d(x, y)$ and $d(\alpha x, \beta x) = |\alpha - \beta| \, \|x\|$, where $\|x\| = d(x, 0)$.

Notice that a retraction set X can be fibred over $0 \cdot X$, with each fibre a star set. We could also define a metric retraction space by putting infinite distance between fibres.

Example 4.2. The real half line $\mathbf{R}^+$ is naturally a topological and metric star space. The interval $[0, 1]$ (or $[0, r]$ for any $r > 0$) is a compact topological and bounded metric star space.

Example 4.3. If X and Y are two star sets, then $X \times Y$, equipped with the diagonal action $\alpha(x, y) = (\alpha x, \alpha y)$, is again a star set. If both are metric star spaces, then equipping the product with the maximum distance makes it a metric star space as well (here the maximum distance is preferable to the sum distance, since it preserves bound hypotheses on the diameter).

Example 4.4. Let X be a set, and equip $[0, 1] \times X$ with the equivalence relation

$$(\alpha, x) \sim (\beta, y) \quad \Longleftrightarrow \quad (\alpha, x) = (\beta, y) \text{ or } \alpha = \beta = 0.$$

The *cone* of X is the quotient space

$$*X = ([0, 1] \times X)/{\sim}.$$

A member of $*X$ is denoted by $[\alpha, x]$. We equip it with the action $\alpha \cdot [\beta, x] = [\alpha\beta, x]$. This makes it a star set, with $[0, x] = 0$ regardless of x.

We tend to identify $x \in X$ with $[1, x] \in *X$, so $[\alpha, x]$ may also be denoted by αx.

When X is a compact Hausdorff space, the relation $\sim$ is closed, $*X$ is again compact and Hausdorff, and the identification $X \subseteq *X$ is a topological embedding. When X is a bounded metric space, say $\operatorname{diam}(X) \leq 2$, we propose to metrise $*X$ by

$$d(\alpha x, \beta y) = |\alpha - \beta| + (\alpha \wedge \beta) d(x, y). \tag{2}$$

In particular, if either α or β vanishes, then the right-hand side does not depend on either x or y, so d is well defined, and $d(0, x) = 1$ for all $x \in X$.

The only property that is not entirely obvious is the triangle inequality, namely

$$|\alpha - \gamma| + (\alpha \wedge \gamma) d(x, z) \leq |\alpha - \beta| + (\alpha \wedge \beta) d(x, y) + |\beta - \gamma| + (\beta \wedge \gamma) d(y, z). \tag{3}$$

We may assume that $\alpha \geq \gamma$, so $\alpha \wedge \gamma = \gamma$. If $\beta \geq \gamma$, then (3) holds trivially since $\alpha \wedge \beta \geq \gamma = \beta \wedge \gamma$. If $\beta \leq \gamma$, then the right-hand side evaluates to

$$(\alpha - \gamma) + 2(\gamma - \beta) + \beta d(x, y) + \beta d(y, z).$$

Applying the triangle inequality for X and the hypothesis that $2 \geq d(x, z)$, we obtain (3) in this case as well.

We conclude that $(*X, d)$ is a metric space. The embedding $X \subseteq *X$ is isometric, and $\operatorname{diam}(*X) = 1 \vee \operatorname{diam}(X)$. If X is complete, then so is $*X$.

A special instance of this is the cone of a singleton, which can be identified with the interval $[0, 1]$ equipped with the natural star, topological or metric structures.

Example 4.5. More generally, let S be a star set, X an arbitrary set, and define

$$(s, x) \sim (t, y) \quad \Longleftrightarrow \quad (s, x) = (t, y) \text{ or } s = t = 0,$$
$$S * X = (S \times X)/{\sim}.$$

As in the definition of a cone, a member of $S * X$ is denoted by $[s, x]$ or $s * x$ (in analogy with the notation αx). We make $S * X$ into a star set by defining $\alpha \cdot (s * x) = (\alpha s) * x$.

This indeed generalises the cone construction, with $*X = [0, 1] * X$.

When S and X are compact Hausdorff spaces, the relation $\sim$ is closed, and $S * X$ is again compact and Hausdorff. When S and X are bounded metric spaces, say $\operatorname{diam}(X) \leq 2$ and $\|s\| \leq 1$ for all $s \in S$, we equip $S * X$ with the distance function

$$d(s * x, t * y) = d(s, t) \vee d(\|s\|x, \|t\|y),$$

where $d(\|s\|x, \|t\|y)$ is calculated in $*X$. Notice that $\|s*x\| = \|s\|$, and the distance functions on $[0, 1] * X$ and $*X$ agree.

Remark 4.6. The generalised cone construction of Example 4.5 can be easily iterated: $S * (X \times Y) = (S * X) * Y$, identifying $s * (x, y) = s * x * y$. In the metric case, assume that X and Y are both of diameter at most two. Equipping products with the maximum distance, $\operatorname{diam}(X \times Y) \leq 2$ as well, and the obvious map $*(X \times Y) \to *X \times *Y$ sending $\alpha(x, y) \mapsto (\alpha x, \alpha y)$ is isometric. It follows that the identification $S \times (X \times Y) = (S * X) * Y$ is isometric:

$$\begin{aligned} d(s * x * y, t * u * v) &= d(s * x, t * u) \vee d(\|s * x\|y, \|t * u\|v) \\ &= d(s, t) \vee d(\|s\|x, \|t\|u) \vee d(\|s\|y, \|t\|v) \\ &= d(s, t) \vee d\big(\|s\|(x, y), \|t\|(u, v)\big) \\ &= d(s * (x, y), t * (u, v)). \end{aligned}$$

In particular, $*(X \times Y) = (*X) * Y$.

Definition 4.7. Let X and Y be two retraction (star) spaces. A map $f : X \to Y$ is *homogeneous* if $f(\alpha x) = \alpha f(x)$. It is *subhomogeneous* if $f(\alpha x) = \beta f(x)$ for some $\beta \leq \alpha$. The latter is mostly used when $Y = \mathbf{R}^+$, in which subhomogeneity becomes $f(\alpha x) \leq \alpha f(x)$.

We may also equip a retraction space with a partial order defined by $\alpha x \leq x$ whenever $\alpha \in [0, 1]$. This induces the usual partial order on $\mathbf{R}^+$, and subhomogeneity can be stated as $f(\alpha x) \leq \alpha f(x)$ for arbitrary maps between retraction spaces. Notice also that our definition of a metric retraction space X simply requires the distance function to be subhomogeneous on $X \times X$.

5. Star sorts

Definition 5.1. A *star sort* is a sort equipped with a definable structure of a metric star space. In particular, this means that the map $(\alpha, x) \mapsto \alpha x$ is definable (and not just $x \mapsto \alpha x$ for each α). Star sorts are usually denoted by D^*, E^*, and so on.

Definition 5.2. Let D^* be a star sort and $\varphi(u, y)$ a formula on $D^* \times E$. We say that φ is *subhomogeneous* if it satisfies $\alpha\varphi(u, y) \geq \varphi(\alpha u, y) \geq 0$.

We may specify that it is subhomogeneous in the variable u, especially if u is not the first variable. More generally, we may say that $\varphi(u, v, \ldots)$ is subhomogeneous in (u, v) if $\alpha\varphi(u, v, \ldots) \geq \varphi(\alpha u, \alpha v, \ldots) \geq 0$, and similarly for any other tuple of variables. If it is subhomogeneous in the tuple of all its variables, we just say that φ is *jointly subhomogeneous*.

Example 5.3. (i) If D is any sort (of diameter at most two), then the cone $*D$, equipped with the distance proposed in Example 4.4, is a star sort. More generally, if D^* is a star sort and E an arbitrary sort, then $D^* * E$, as per Example 4.5, is a star sort.

(ii) Any finite product of star sorts, equipped with the diagonal action of $[0, 1]$ and the maximum or sum distance, is again a star sort. Similarly, any countable product of star sorts, equipped with

$$d(u, v) = \sum_n \frac{d_n(u_n, v_n)}{2^n \operatorname{diam}(d_n)},$$

is again a star sort, and the same holds with supremum in place of sum.

(iii) If D^* is a star sort and $d'(u, v)$ a jointly subhomogeneous definable pseudodistance on D^*, then the quotient (D^*, d') can be equipped with an induced star structure, making it again a star sort.

(iv) Let D^* be a star sort and $E^* \subseteq D^*$ a definable subset. Then the distance $d(u, E^*)$ is subhomogeneous if and only if E^* is closed under multiplication by $\alpha \in [0, 1]$, in which case E^* is again a star sort.

Notice that $\varphi(u, y)$ is subhomogeneous in u if for every fixed parameter b, the formula $\varphi(u, b)$ (in u alone) is subhomogeneous.

For an alternate point of view, notice that a subhomogeneous formula $\varphi(u, y)$ does not depend on y when $u = 0$. It can therefore be viewed as a formula $\varphi(u * y)$ in the sort $D^* * E$ (see Example 4.5). Since $\alpha(u * y) = (\alpha u) * y$, a subhomogeneous (in u) formula $\varphi(u, y)$ is the same thing as a subhomogeneous formula $\varphi(u * y)$ in a single variable from the sort $D^* * E$.

Similarly, a formula $\varphi(u, v)$ on $D^* \times E^*$ is jointly subhomogeneous if and only if it is subhomogeneous as a formula on the product star sort.

Question 5.4. We ordered the clauses of Example 5.3 in order to reflect the three operations by which we construct sorts in general. Still, something more probably needs to be said regarding the construction of subhomogeneous pseudodistance functions. In the usual context of plain sorts (and plain pseudodistances), to every formula $\varphi(x, t)$ on $D \times E$ we can associate a formula on $D \times D$, defined by

$$d_\varphi(x, y) = \sup_t |\varphi(x, t) - \varphi(y, t)|.$$

This is always a definable pseudodistance on D. Moreover, in the case where $E = D$ and φ already defines a pseudodistance, d_φ agrees with φ.

Can something analogous be done in the present context as well?

The following essentially asserts that we can retract continuously (with Lipschitz constant one, even) all formulas into subhomogeneous ones. The analogous result for a formula in several variables, with respect to joint subhomogeneity in some of them, follows.

Proposition 5.5. *Let D^* be a star sort and $\varphi(u, y) \geq 0$ a positive formula on $D^* \times E$. For $k \in \mathbf{N}$, define*

$$(\mathrm{SH}_k\, \varphi)(u, y) = \inf_{u',\alpha} (\alpha\varphi(u', y) + kd(\alpha u', u)), \quad \text{where } u' \in D^*,\ \alpha \in [0, 1].$$

(i) *For any $\varphi \geq 0$ and k, the formula $(\mathrm{SH}_k\, \varphi)(u, y)$ is k-Lipschitz and subhomogeneous in u, and $\mathrm{SH}_k\, \varphi \leq \varphi$.*

(ii) *For any two formulas $\varphi, \psi \geq 0$ and $r \geq 0$, if $\varphi \leq \psi + r$, then $\mathrm{SH}_k\, \varphi \leq (\mathrm{SH}_k\, \psi) + r$. Consequently, $|(\mathrm{SH}_k\, \varphi) - (\mathrm{SH}_k\, \psi)| \leq |\varphi - \psi|$.*

(iii) *If φ is subhomogeneous, then $(\mathrm{SH}_k\, \varphi) \to \varphi$ uniformly, at a rate that only depends on the bound and uniform continuity modulus of φ.*

Proof. Clearly, $(\mathrm{SH}_k\, \varphi)(u, y)$ is k-Lipschitz in u. If $(\mathrm{SH}_k\, \varphi)(u, y) < r$ and $\beta \in [0, 1]$, then there exist u' and α such that $\alpha\varphi(u', y) + kd(\alpha u', u) < r$. Then

$$\alpha\beta\varphi(u', y) + kd(\alpha\beta u', \beta u) < \beta r,$$

showing that $(\mathrm{SH}_k\, \varphi)(\beta u) < \beta r$. This proves subhomogeneity. We also always have $(\mathrm{SH}_k\, \varphi)(u, y) \leq 1 \cdot \varphi(u, y) + d(1 \cdot u, u) = \varphi(u, y)$.

The second item is immediate.

For the third item, we assume that φ is subhomogeneous, in which case

$$(\mathrm{SH}_k\, \varphi)(u, y) = \inf_{u'} (\varphi(u', y) + kd(u', u)) \leq \varphi(u, y).$$

Say that $|\varphi| \leq M$ and $d(u, u') < \delta$ implies $|\varphi(u, y) - \varphi(u', y)| < \varepsilon$, and let $k > 2M/\delta$. If $d(u', u) \geq \delta$, then $\varphi(u', y) + kd(u', u) \geq \varphi(u)$, so such u' may be ignored. Restricting to those where $d(u', u) < \delta$, we see that $(\mathrm{SH}_k\, \varphi) \geq \varphi - \varepsilon$. □

Definition 5.6. We say that a formula $\varphi(x, y)$ is *witness-normalised* (in x, unless another variable is specified explicitly) if $\inf_y \varphi = 0$ (equivalently, if $\varphi \geq 0$ and $\sup_x \inf_y \varphi = 0$).

More generally, for $\varepsilon > 0$, we say that $\varphi(x, y)$ is *ε-witness-normalised* (in x) if $0 \leq \inf_y \varphi \leq \varepsilon$.

Witness-normalised formulas are analogous to formulas $\varphi(x, y)$ in classical logic for which $\exists y\, \varphi$ is valid: in either case, we require that witnesses exist. If $\varphi(x, y)$ is any formula, then $\varphi(x, y) - \inf_z \varphi(x, z)$ is witness-normalised (we may say that it is *syntactically* witness normalised), where we subtract a "normalising" term.

By definition, a subhomogeneous or a witness-normalised formula is positive. If φ is witness-normalised in any of its arguments and $\varphi \geq \psi \geq 0$, then so is ψ. This applies in particular to the formulas $\mathrm{SH}_k\,\varphi$ constructed in Proposition 5.5, assuming φ is witness-normalised.

Definition 5.7. Let D^* and E^* be two star sorts. A *star correspondence* between D^* and E^* is a formula $\varphi(u, v)$ on $D^* \times E^*$ that is subhomogeneous in (u, v) and witness-normalised in each of u and v.

Similarly, an ε-*star correspondence* is a jointly subhomogeneous formula that is ε-witness-normalised in each argument.

Remark 5.8. If φ is ε-witness-normalised (in one of its variables), then $\varphi' = \varphi \dot{-} \varepsilon$ is witness-normalised (in the same), and $|\varphi - \varphi'| \leq \varepsilon$. If φ is subhomogeneous, then so is $\varphi \dot{-} \varepsilon$,

Therefore, if φ is an ε-star correspondence, then $\varphi' = \varphi \dot{-} \varepsilon$ is a star correspondence, and $|\varphi - \varphi'| \leq \varepsilon$.

Say that a definable map $\sigma : D \to E$ is *densely surjective* if it is surjective in every sufficiently saturated model of the ambient theory, or equivalently, if σ has dense image in every model. Recall that a definable map $\sigma : D^* \to E^*$ between star sorts is *homogeneous* if $\sigma(\alpha u) = \alpha\sigma(u)$.

Notice that a definable map $\sigma : D^* \to E^*$ is homogeneous if and only if the formula $d(\sigma u, v)$ is subhomogeneous in (u, v), and it is always witness-normalised in u. If σ is densely surjective, then it is homogeneous if and only if $d(\sigma u, v)$ is a star correspondence. If σ is bijective, then this is further equivalent to $d(u, \sigma^{-1} v)$ being a star correspondence.

Definition 5.9. Say that a star sort D^* is *universal* (as a star sort) if for every star sort E^*, every star correspondence φ between D^* and E^*, and every $\varepsilon > 0$, there exists a $\frac{1}{2}$-star correspondence ψ such that, in addition, if $\psi(u, v_i) < 1$ for $i = 0, 1$, then $\varphi(u, v_i) < \varepsilon$ and $d(v_0, v_1) < \varepsilon$.

This just says that condition (ii) of Proposition 5.10, which may be easier to parse, holds "approximately". The choice of 1 and $\frac{1}{2}$ is quite arbitrary, and any two constants $0 < r_1 < r_2$ would do just as well (in the proof of Proposition 5.10(i) below, replace $2\psi \dot{-} 1$ with $(\psi \dot{-} r_1)/(r_2 - r_1)$).

Proposition 5.10. *Let D^* and E^* be star sorts, $\varphi(u, v)$ a star correspondence on $D^* \times E^*$, and $\varepsilon > 0$.*

(i) *If D^* is a universal star sort, then there exists ψ as in Definition 5.9 that is a star correspondence (rather than a mere $\frac{1}{2}$-star correspondence).*

(ii) *If D^* is a universal star sort, then there exists a densely surjective homogeneous definable map $\sigma : D^* \to E^*$ such that $\varphi(u, \sigma u) \leq \varepsilon$.*

(iii) *If both D^* and E^* are both universal star sorts, then the same can be achieved with σ bijective.*

Proof. For (i), let ψ be as in the conclusion of Definition 5.9. Then $2\psi \dot{-} 1$ will do.

For (ii), define a sequence of formulas $\varphi_n(u, v)$ as follows. We start with $\varphi_0 = \varphi$, and we may assume that $0 < \varepsilon < 1$. Then, assuming that φ_n is a star correspondence, we find a star correspondence φ_{n+1} such that $\varphi_{n+1}(u, v_i) < 1$ implies $\varphi_n(u, v_i) \leq \varepsilon$ and $d(v_0, v_1) < \varepsilon/2^n$. Let $X_n \subseteq D^* \times E^*$ be the (type-definable) set defined by $\varphi_n \leq \varepsilon$ and $X = \bigcap X_n$. By hypothesis, for every $u \in D^*$ and n, there exists $v \in E^*$ such that $(u, v) \in X_n$. We also have $X_{n+1} \subseteq X_n$, so in a sufficiently saturated model there exists $v \in E^*$ such that $(u, v) \in X$. By the second hypothesis on φ_n, such v is unique, so X is the graph of a definable map σ (and v belongs to any model that contains u). By the same reasoning as above, for every $v \in E^*$ there exists $u \in D^*$ (not necessarily unique, so potentially only in a sufficiently saturated model) such that $(u, v) \in X$, so σ is densely surjective.

Assume now that $v = \sigma u$, i.e., $(u, v) \in X$. Since each φ_n is subhomogeneous, $(\alpha u, \alpha v) \in X$ for every $\alpha \in [0, 1]$, i.e., $\alpha v = \sigma(\alpha u)$, and σ is homogeneous. Finally, since $\varphi_0 = \varphi$, we have $(u, \sigma u) \in X \subseteq X_0$, so $\varphi(u, \sigma u) \leq \varepsilon$.

For (iii) we use a back-and-forth version of the previous argument, with the roles of D^* and E^* reversed at odd steps. □

Notice that the zero formula is (trivially) a star correspondence on any two star sorts. Therefore, if a universal star sort exists, then it is unique, up to a homogeneous definable bijection.

Lemma 5.11. *Let (D^*_n) be an inverse system of star sorts, where each $\pi_n : D^*_{n+1} \to D^*_n$ is surjective and homogeneous.*

(i) *The inverse limit $D^* = \varprojlim D^*_n$ is a star sort, with the natural action $\alpha(u_n) = (\alpha u_n)$ and the distance proposed in Example 5.3.*

(ii) *A star correspondence between D^* and E^* that factors through $D^*_n \times E^*$ is the same thing as a star correspondence between D^*_n and E^*.*

(iii) *In order for D^* to be a universal star sort, it is enough for it to satisfy the condition of Definition 5.9 for star-correspondences φ that factor through $D^*_n \times E^*$ for some n.*

Proof. The first two assertions are fairly evident. In what follows, we are going to identify a formula $\varphi(u_n, v)$ on $D^*_n \times E^*$ with the formula $\varphi(\pi_n(u), v)$ on $D^* \times E^*$, which is essentially what the second point says.

For the last one, say that φ is a star correspondence between D^* and E^*, and let $\varepsilon > 0$. For n large enough we may find a formula $\varphi_1(u_n, v)$ on $D^*_n \times E^*$ such that $\varphi \geq \varphi_1 \geq \varphi \dot{-} \varepsilon$ (with the identification proposed in the previous paragraph). Since φ is jointly subhomogeneous, so is $\varphi \dot{-} \varepsilon$. Using the construction of Proposition 5.5, this implies that for large enough k we have

$$\varphi \geq \mathrm{SH}_k\,\varphi \geq \mathrm{SH}_k\,\varphi_1 \geq \mathrm{SH}_k(\varphi \dot{-} \varepsilon) \geq \varphi \dot{-} 2\varepsilon.$$

Since $\varphi' = \mathrm{SH}_k\,\varphi_1$ is jointly subhomogeneous, it is a star correspondence, and it factors through $D_n^* \times E^*$. Assume now that $\psi(u, v)$ exists, as per Definition 5.9, for φ' and ε. In particular, if $\psi(u, v) < 1$, then $\varphi'(u, v) < \varepsilon$, so $\varphi(u, v) < 3\varepsilon$, which is good enough. □

6. Sorts with witnesses

In this section, we provide an explicit construction of a universal star sort. We follow a path similar to the construction of D_Φ in [Ben Yaacov 2022], seeking a sort that contains "all witnesses".

Let us consider first the case of a single formula $\varphi(x, y)$ on $D \times E$, which we assume to be witness-normalised (namely, such that $\inf_y \varphi = 0$; see Definition 5.6). The sort D is viewed as the sort of *parameters*, and E is the sort of potential *witnesses*. One may then wish to consider the set of "parameters with witnesses", namely the collection of all pairs (x, y) such that $\varphi(x, y) = 0$, but this may be problematic for several reasons.

First of all, in a fixed (nonsaturated) structure, for all a there exist b such that $\varphi(a, b)$ is arbitrarily small, but not necessarily such that $\varphi(a, b) = 0$. This can be overcome by allowing an error, e.g., by considering all the solution set of $\varphi(x, y) \leq \varepsilon$ for some $\varepsilon > 0$. In fact, it is enough to consider the solution set of $\varphi(x, y) \leq 1$: if we want a smaller error, we need only replace φ with φ/ε.

A second, and more serious issue, is that the resulting set(s) need not be definable. That is to say that it may happen that $1 < \varphi(a, b) < 1 + \varepsilon$ for arbitrarily small $\varepsilon > 0$ without there existing a pair (a', b') close to (a, b) such that $\varphi(a', b') \leq 1$. We can solve this by allowing a *variable error*, considering triplets (r, x, y) where $r \in \mathbf{R}$ and $\varphi(x, y) \leq r$. Now, if $\varphi(x, y) < r + \varepsilon$, then the triplet (r, x, y) is very close to $(r + \varepsilon, x, y)$, which does belong to our set.

This may seem too easy, and raises some new issues. For example, if we allow errors greater than the bound for φ, then the condition $\varphi(x, y) \leq r$ becomes vacuous. This is not, in fact, a real problem, since soon enough we are going to let φ vary (or more precisely, consider an infinite family of formulas simultaneously), and any finite bound r will be meaningful for *some* of the formulas under consideration. However, in order for the previous argument to work, r cannot be bounded (we must always be able to replace it with $r + \varepsilon$). By compactness, $r = +\infty$ must be allowed as well — and now there is no way around the fact that $\varphi(x, y) \leq \infty$ is vacuous, regardless of φ.

We seem to be chasing our own tail, each time shovelling the difficulty underneath a different rug — indeed, a complete solution is impossible, or else we could construct a universal Skolem sort, which was shown in [Ben Yaacov 2022] to be impossible in general. What we propose here is a "second best": allow infinite error, but use the formalism of star sorts to identify all instances with infinite error as the distinguished root element. Thus, at the root, all information regarding the (meaningless) witnesses will be lost, while every point outside the root will involve finite error, and therefore meaningful witnesses. Since we want the root to be at zero, rather than at infinity, we replace $r \in [1, \infty]$ with $\alpha = 1/r \in [0, 1]$.

Let D^* be a star sort, E a sort. The set $D^* * E = \{u * y : u \in D^*,\ y \in E\}$, as per Example 4.5, is again a star sort, in which $0 * y = 0$ regardless of y.

Lemma 6.1. *Let D^* be a star sort, E a sort, and $\varphi(u, y)$ a formula on $D^* \times E$, witness-normalised and subhomogeneous in u. Then*

$$D^*_\varphi = \{u * y : u \in D^* \text{ and } \varphi(u, y) \leq 1\} \subseteq D^* * E$$

is again a star sort, and the natural projection map $D^*_\varphi \to D^*$, sending $u * y \mapsto u$, is surjective.

Proof. We may view φ as a formula on $D^* * E$, since, by subhomogeneity, $\varphi(0, y) = 0$ regardless of y. The set D^*_φ is the zero-set in $D^* * E$ of the formula $\varphi \dot{-} 1$. Assume now that $a * b \in D^* * E$ and $\varphi(a, b) \dot{-} 1 < \delta$. Then $(1 - \delta)a * b \in D^*_\varphi$, and it is as close as desired (given δ small enough) to $a * b$. Therefore, D^*_φ is definable. Since φ is subhomogeneous, D^*_φ is closed under multiplication by $\alpha \in [0, 1]$ and is therefore a star sort. Since φ is witness-normalised, the projection is onto. □

Let us iterate this construction. Recall from Remark 4.6 that $(*D) * E = *(D \times E)$, identifying $(\alpha x) * y = \alpha(x, y)$. Therefore, if $D^* \subseteq *D$ (with the induced star structure), then $D^* * E \subseteq *(D \times E)$.

Definition 6.2. Fix a sort D, as well as a sequence of formulas $\Phi = (\varphi_n)$, where each $\varphi_n(x_{<n}, y)$ is a witness-normalised formula on $D^n \times D$. Since Φ determines the sort D, we say that Φ is a sequence *on* D. We then define

$$D^*_n = \{\alpha x_{<n} : \alpha\varphi_k(x_{<k}, x_k) \leq 1 \text{ for all } k < n\} \subseteq *(D^n),$$
$$D^*_\Phi = \{\alpha x : \alpha\varphi_n(x_{<n}, x_n) \leq 1 \text{ for all } n\} \subseteq *(D^{\mathbf{N}}).$$

In other words,

$$D^*_0 = [0, 1] = *(\text{singleton}), \qquad D^*_{n+1} = (D^*_n)_{\varphi'_n}, \qquad D^*_\Phi = \varprojlim D^*_n,$$

where $\varphi'_n(\alpha x_{<n}, y) = \alpha\varphi(x_{<n}, y)$. By Lemma 6.1, each D^*_n is a star sort, and the natural projection $D^*_{n+1} \to D^*_n$ is onto. By Lemma 1.5, $D^*_\Phi = \varprojlim D^*_n$ is also a sort, and therefore a star sort by Lemma 5.11.

Notice that any formula in D^*_n can be viewed, implicitly, as a formula in D^*_k for any $k \geq n$, or even in D^*_Φ, via the projections $D^*_k \twoheadrightarrow D^*_n$ or $D^*_\Phi \twoheadrightarrow D^*_n$ (this is, essentially, an addition of dummy variables). In what follows, variables in D^*_n are denoted by u_n or $\alpha x_{<n}$ (where $x_{<n} \in D^n$), and similarly, variables in D^*_Φ are denoted by u or αx.

Definition 6.3. We say that the sequence Φ on a sort D is *rich* if D admits a definable projection onto any countable product of basic sorts, and for every witness-normalised formula $\varphi(x_{<n}, y)$ in $D^n \times D$ and every $\varepsilon > 0$ there exist arbitrarily big $k \geq n$ such that $|\varphi_k(x_{<k}, y) - \varphi(x_{<n}, y)| < \varepsilon$ (so φ is viewed as a formula in $x_{<k}, y$ through the addition of dummy variables).

Lemma 6.4. *Under our standing hypothesis that the language is countable, with countably many basic sorts, there exists a rich sequence Φ (on an appropriate sort D). Moreover, we may construct Φ (and D) in a manner that only depends on the language and not on the theory of any specific structure.*

Proof. For D we may take the (countable) product of all infinite countable powers of the basic sorts. For each k we may choose a countable dense family of formulas on $D^k \times D$, call them $\psi_{k,m}(x_{<k}, y)$. Replacing them with $\chi_{k,m}(x_{<k}, y) = \psi_{k,m}(x_{<k}, y) - \inf_z \psi_{k,m}(x_{<k}, z)$, we obtain a countable dense family of witness-normalised (in $x_{<k}$) formulas on $D^k \times D$. We may now construct a rich sequence Φ in which each $\chi_{k,m}$ occurs infinitely often (with additional dummy x variables). $\square$

Let $\Phi = (\varphi_n)$ (and D) be fixed, with Φ rich. We define a formula on D^n by

$$\rho_n(x_{<n}) = \frac{1}{1 \vee \bigvee_{k<n} \varphi_k(x_{<k}, x_k)}.$$

In other words, $\rho_n(x_{<n})$ is the maximal $\alpha \in [0, 1]$ such that $\alpha x_{<n} \in D_n^*$, or equivalently, such that $x_{<n}$ can be extended to x with $\alpha x \in D_\Phi^*$.

Lemma 6.5. *Let $\Phi = (\varphi_n)$ be rich. Let E^* be another star sort, $\psi(u_n, v)$ a star correspondence on $D_\Phi^* \times E^*$ that factors through $D_n^* \times E^*$, and $\varepsilon > 0$. Then ψ factors through $D_k^* \times E^*$ for every $k \geq n$, and for every large enough k the formula $\psi_1^k(x_{<k}, v) = \psi(\rho_k(x_{<k})x_{<n}, v)$ is ε-witness-normalised in either argument.*

Proof. If $k \geq n$, then $\rho_k(x_{<k}) \leq \rho_n(x_{<n})$, so $\rho_k(x_{<k})x_{<n} \in D_n^*$. Since $\psi(u_n, v)$ is witness-normalised in u_n, $\psi_1^k(x_{<k}, v)$ is witness-normalised in $x_{<k}$. It is left to show that for k large enough, it is also ε-witness-normalised in v.

Our hypothesis regarding D implies, among other things, that there exists a surjective definable map $\chi : D \to [0, 1]$ (namely, a surjective formula). Therefore, for a constant C that we shall choose later, there exists $m \geq n$ such that

$$C\chi(y) \geq \varphi_m(x_{<m}, y) \geq C\chi(y) - 1/C.$$

Assume that $k > m$. For every possible value of $v \in E^*$, which we consider as fixed, there exists $\alpha x_{<n} \in D_n^*$ such that $\psi(\alpha x_{<n}, v) < \varepsilon$. We can always extend $x_{<n}$ to $x_{<m}$ in such a manner that $\rho_m(x_{<m}) = \rho_n(x_{<n}) \geq \alpha$, so $\alpha x_{<m} \in D_m^*$. We choose x_m so $\chi(x_m) = (\alpha C \vee 1)^{-1}$, and extend $x_{\leq m}$ to $x_{<k}$ so $\rho_k(x_{<k}) = \rho_{m+1}(x_{\leq m})$.

If $\alpha C \geq 1$, then $1/\alpha \geq \varphi_m(x_{<m}, x_m) \geq 1/\alpha - 1/C$, so

$$\alpha \leq \rho_{m+1}(x_{\leq m}) \leq \alpha(1-\alpha/C)^{-1}.$$

Having chosen C large enough, $\rho_k(x_{<k}) = \rho_{m+1}(x_{\leq m})$ is as close to α as desired. If $\alpha C < 1$, then $0 \leq \alpha \leq 1/C$ and $0 < \rho_{k+1}(x_{\leq k}) \leq 1/(C-1/C)$, so the same conclusion holds. Either way, having chosen C large enough, $\psi_1^k(x_{<k}, v)$ is as close as desired to $\psi(\alpha x_{<n}, v)$, and in particular $\psi_1^k(x_{<k}, v) < 2\varepsilon$, which is good enough. $\square$

Given our hypothesis regarding D, every sort can be expressed as a definable subset of a quotient of D by a pseudodistance. Such a quotient is denoted by $(D, \bar{d})$ (which includes an implicit step of identifying points at $\bar{d}$-distance zero).

Convention 6.6. From this point, and through the proof of Lemma 6.8, we fix a star sort E^*. By the preceding remark, we may assume that $(E^*, d_{E^*}) \subseteq (D, \bar{d})$ isometrically, where $\bar{d}$ is a definable pseudodistance on D, which we also fix. In particular, the distance on E^* is also denoted by $\bar{d}$. If $y \in D$, we denote its image in the quotient $(D, \bar{d})$ by $\bar{y}$.

It is worthwhile to point out that if $\alpha x \in D^*_\Phi$, then for every $k \in \mathbf{N}$ and $\delta > 0$,

$$(\alpha\delta/2)(\varphi_k(x_{<k}, x_k) + 1) = (\delta/2)(\alpha\varphi_k(x_{<k}, x_k) + \alpha) \leq \delta. \tag{4}$$

Given $n \leq k$ and $\delta > 0$, let us define, for $\alpha x \in D^*_\Phi$, $v \in E^*$ and $y \in D$,

$$\chi^n(\alpha x, y, v) = \inf_{w \in E^*} \bigl[\bar{d}(\alpha\rho_n(x_{<n})^{-1}w, v) + \alpha\bar{d}(\bar{y}, w)\bigr],$$

$$\chi^{n,k}(\alpha x, v) = \chi^n(\alpha x, x_k, v) = \inf_{w \in E^*} \bigl[\bar{d}(\alpha\rho_n(x_{<n})^{-1}w, v) + \alpha\bar{d}(\overline{x_k}, w)\bigr].$$

Let us explain this. First of all, since $\alpha x \in D^*_\Phi$, we must have $\alpha \leq \rho_n(x_{<n})$, so the expression $\alpha\rho_n(x_{<n})^{-1}w$ makes sense. Also, if $\alpha = 0$, then $\chi^n(\alpha x, y, v) = \|v\|$ does not depend on x, so this is well defined.

Now, let $y \in D$ (possibly, $y = x_k$ for some $k \geq n$, but this will happen later). We want v to be equal to $\alpha\rho_n(x_{<n})^{-1}\bar{y}$, and in particular, we want $\bar{y}$ to belong to E^*. We may not multiply by $\alpha\rho_n(x_{<n})^{-1}$ outside E^*, but we may quantify over E^*. Therefore, we ask for $\bar{y}$ to be very close to some $w \in E^*$, and for $\alpha\rho_n(x_{<n})^{-1}w$, which always makes sense, to be close to v.

Lemma 6.7. *The formula $\chi^{n,k}(u, v)$ has the following properties*:

(i) *It is jointly subhomogeneous in its arguments.*

(ii) *For every n, $\varepsilon > 0$ there exists $\delta = \delta(n, \varepsilon) > 0$ such that, if $\chi^n(u, y, v_i) \leq \delta$ for $i = 0, 1$, then $\bar{d}(v_0, v_1) < \varepsilon$. In particular, for any k, if $\chi^{n,k}(u, v_i) \leq \delta$ for $i = 0, 1$, then $\bar{d}(v_0, v_1) < \varepsilon$.*

(iii) *Assuming that $\varphi_k(x_{<k}, y) \geq 2\bar{d}(\bar{y}, E^*)/\delta - 1$, the formula $\chi^{n,k}(u, v)$ is δ-witness-normalised in u.*

Proof. Item (i) is immediate (among other things, we use the fact that $\bar{d}$ is subhomogeneous on E^*).

For (ii), assume that $\chi^n(\alpha x, y, v_i) = 0$. Then either $\alpha = 0$, in which case $v_i = 0$, or $\alpha > 0$, in which case we have $\bar{y} \in E^*$ and $v_i = \alpha\rho_n(x_{<n})^{-1}\bar{y}$. Either way, $v_0 = v_1$, and in particular $\bar{d}(v_0, v_1) < \varepsilon$. The conclusion follows by compactness.

For (iii), let $u = \alpha x \in D^*_\Phi$. By (4) we have $\alpha\bar{d}(\overline{x_k}, E^*) \le \delta$. Choose $w \in E^*$ such that $\alpha\bar{d}(\overline{x_k}, w) \le \delta$, and let $v = \alpha\rho_n(x_{<n})^{-1}w$. Then $\chi^{n,k}(u, v) \le \delta$. □

Lemma 6.8. *Let $\Phi = (\varphi_n)$ be rich. Let $E^* \subseteq (D, \bar{d})$ be a star sort, as per Convention 6.6, $\psi(u, v)$ a star correspondence on $D^*_\Phi \times E^*$, and $\varepsilon > 0$. Then there exist $n \le k$ and $\delta > 0$ such that $\chi^{n,k}(u, v)$ is a δ-star correspondence between D^*_Φ and E^*, and in addition, if $\chi^{n,k}(u, v_i) \le 2\delta$ for $i = 0, 1$, then $\psi(u, v_i) \le \varepsilon$ and $\bar{d}(v_0, v_i) < \varepsilon$.*

Proof. By Lemma 5.11 and Lemma 6.5, for some n (in fact, any n large enough), we may assume that ψ is a star correspondence that factors as $\psi(u_n, v)$ through $D^*_n \times E^*$, and that $\psi_1(x_{<n}, v) = \psi(\rho_n(x_{<n})x_{<n}, v)$ is ε-witness-normalised in either argument. In particular, $\psi_1 \dot{-} \varepsilon$ is witness-normalised.

We may extend $\psi_1 \dot{-} \varepsilon$ to $D^n \times (D, \bar{d})$, obtaining a formula $\psi_2(x_{<n}, y)$ on $D^n \times D$, which is uniformly $\bar{d}$-continuous in y. Since $\psi_1 \ge 0$, we may assume that $\psi_2 \ge 0$, and even that

$$\psi_2(x_{<n}, y) \ge \bar{d}(\bar{y}, E^*).$$

Let us choose $\delta > 0$ small enough, based on choices made so far. Since $\psi_2(x_{<n}, y)$ is witness-normalised in $x_{<n}$ (choosing witnesses $\bar{y} \in E^*$), there exists $k \ge n$ such that $|\varphi_k - 2\psi_2/\delta| \le 1$. By Lemma 6.7, having chosen δ small enough, the formula $\chi^{n,k}(u, v)$ is jointly subhomogeneous, δ-witness-normalised in u, and $\chi^{n,k}(u, v_i) \le 2\delta$ implies $\bar{d}(v_0, v_i) < \varepsilon$. Two more properties remain to be checked.

First, we need to check that $\chi^{n,k}(u, v)$ is δ-witness-normalised in v. Indeed, given $v = \bar{y} \in E^*$, we know that there exists a sequence $x_{<n} \in D^n$ such that $\psi_1(x_{<n}, v) \dot{-} \varepsilon = 0$. Let $\alpha = \rho_n(x_{<n})$, so $\alpha x_{<n} \in D^*_n$, and extend the sequence $x_{<n}$ to $x_{<k}$ keeping $\alpha x_{<k} \in D^*_k$. We now choose $x_k = y$, so $\psi_2(x_{<n}, x_k) = 0$ and $\varphi_k(x_{<k}, x_k) \le 1$. Therefore, $\alpha x_{\le k} \in D^*_{k+1}$, and we may complete the sequence to $x \in D^{\mathbf{N}}$ such that $\alpha x \in D^*_\Phi$. Then $\chi^{n,k}(\alpha x, v) = 0$, as witnessed by $w = v$ (recalling that we chose $\alpha = \rho_n(x_{<n})$).

Second, we need to check that, having chosen δ appropriately, $\chi^{n,k}(\alpha x, v) \le 2\delta$ implies $\psi(\alpha x, v) \le \varepsilon$. Indeed, following a path similar to the proof of Lemma 6.7(ii), assume that

$$\chi^n(\alpha x, y, v) = \alpha\psi_2(x_{<n}, y) = 0.$$

If $\alpha = 0$, then $v = 0$ and $\psi(\alpha x, v) = \psi(0, 0) = 0$. If $\alpha > 0$, then $\bar{y} \in E^*$, $v = \alpha\rho_n(x_{<n})^{-1}\bar{y}$, and $\psi(\rho_n(x_{<n})x, \bar{y}) \dot{-} \varepsilon = \psi_2(x_{<n}, y) = 0$. Since $(\alpha x, v) = \alpha\rho_n(x_{<n})^{-1}(\rho_n(x_{<n})x, \bar{y})$, it follows that $\psi(\alpha x, v) \le \varepsilon$ in this case as well. By

compactness, for δ small enough, if $\chi^n(\alpha x, y, v) \leq 2\delta$ and $\alpha\psi_2(x_{<n}, y) \leq \delta$, then $\psi(\alpha x, v) < 2\varepsilon$. This last argument does not depend on k, so we may assume that δ was chosen small enough to begin with. By (4), the inequality $\alpha\psi_2(x_{<n}, x_k) \leq \delta$ is automatic when $\alpha x \in D^*_\Phi$. If, in addition, we assume that $\chi^{n,k}(\alpha x, v) = \chi^n(\alpha x, x_k, v) \leq 2\delta$, then $\psi(\alpha x, v) < \varepsilon$, completing the proof. □

Theorem 6.9. *Let Φ be a rich sequence. Then D^*_Φ is universal.*

Proof. Immediate from Lemma 6.8, using the formula $2\chi^{n,k}/\delta$. □

Let us sum up everything we know about the existence and uniqueness of universal star sorts.

Corollary 6.10. *Every theory T (not necessarily complete) admits a universal star sort, which is unique up to a bijective homogeneous map. Moreover, this unique universal sort only depends on the bi-interpretation class of T.*

For a more precise statement of the moreover part, assume that T and T' are bi-interpretable, so by Definition 1.4, they admit a common interpretational expansion T''. Then any two universal star sorts D^*_T and $D^*_{T'}$ of T and T', respectively, are also universal star sorts of T''. As such, they admit a definable homogeneous bijection.

Proof. For any theory T, the existence is by Theorem 6.9, and the uniqueness by Proposition 5.10(iii).

Let us consider two theories T and T', and assume that T' is an interpretational expansion of T. Let D^* be a star sort of T. Since T' is an expansion of T, D^* is also a star sort of T'. Conversely, if E^* is a star sort of T', then, since T' is an *interpretational* expansion of T, E^* admits a definable bijection (in the sense of T') with a sort of T, call it $\widehat{E}^*$. This definable bijection induces a star sort structure on $\widehat{E}^*$. Since T' is an interpretational expansion of T, it cannot introduce new structure on sorts already interpretable in T. Therefore, the star sort structure on $\widehat{E}^*$ is definable in T. In other words, T and T' have the same star sorts.

Now let D^*_T be a universal star sort of T. Then D^*_T is also a star sort of T'. We have just seen that every star sort of T' is also a star sort of T, so every instance of the condition of Definition 5.9 for D^*_T in T' can be translated to such an instance in T. Therefore, D^*_T is also a universal star sort of T'. The moreover part follows. □

7. Further properties of the universal star sort

In Section 5 we showed that the universal star sort, if it exists, is unique up to a homogeneous definable bijection, and in Section 6 we showed that one exists as D^*_Φ for any rich sequence Φ. Let us prove a few additional properties of this special sort.

Convention 7.1. From now on, D^* denotes any universal star sort. Since it is unique up to a homogeneous definable bijection, multiplication by $\alpha \in [0, 1]$ is well defined regardless of the construction we choose for D^*. In particular, its root is well defined.

Notice that we can construct it as D^*_Φ in a manner that only depends on the language (and not on T): we obtain a universal star sort for T simply by restricting our consideration of this sort to models of T.

The uniqueness of D^* means that we may choose it to be D^*_Φ for any rich Φ, and in particular, that we are allowed some leverage in choosing a convenient sequence Φ, as in the proof of the following result.

Theorem 7.2. *The universal star sort D^* is a coding sort for any theory T* (see Definition 2.4), *with the exceptional set being the root $D^0 = \{0\}$.*

Proof. Being a coding sort (with some exceptional set) is invariant under definable bijections (that preserve the exceptional set). Therefore, despite the fact that D^* is only well defined up to a homogeneous definable bijection, our statement makes sense. We may choose a rich sequence Φ on a sort D, as per Definition 6.3, and take $D^* = D^*_\Phi$.

Let $M \vDash T$ and $\alpha a \in D^*_\Phi(M) \smallsetminus \{0\}$, and let

$$N = \mathrm{dcl}(\alpha a) \subseteq M,$$

necessarily a closed set (if M is multisorted, closed in each sort separately). Then $\alpha \neq 0$, and $N = \mathrm{dcl}(a)$. In order to show that $N \preceq M$, it suffices to show that it satisfies the Tarski–Vaught criterion: for every formula $\varphi(x, y)$, where x is in the sort $D^{\mathbf{N}}$ and y is in one of the basic sorts,

$$\inf_y \varphi(a, y) = \inf_{b \in N} \varphi(a, b),$$

where the truth values are calculated in M. Since D projects, by hypothesis, onto any basic sort, we replace φ with its pull-back and assume that it is a formula on $D^{\mathbf{N}} \times D$. Replacing φ with $\varphi(x, y) - \inf_z \varphi(x, z)$, we may assume that φ is witness-normalised and the left-hand side vanishes. Then it is enough to show that for every $\varepsilon > 0$ there exists $b \in N$ such that $\varphi(a, b) < \varepsilon$, and replacing φ with an appropriate multiple, it is enough to require $\varphi(a, b) \leq 1 + 1/\alpha$. Choosing n such that φ_n is a good-enough approximation of φ, it is enough to find $b \in D(N)$ such that $\varphi_n(a_{<n}, b) \leq 1/\alpha$. For this, $b = a_n$ will do. This proves the coding models property of Definition 2.4.

For the density property, assume that M is separable, and let $\alpha a \in D(M)$. Assume first that $\alpha > 0$. We may freely assume that $\varphi_k = 0$ infinitely often. Let us fix n_0, and define a sequence $b \in D^{\mathbf{N}}$ as follows.

- We start with $b_{<n_0} = a_{<n_0}$.
- Having chosen $b_{<k}$ (for $k \geq n_0$) such that $\alpha b_{<k} \in D^*_k$, we can always choose $b_k \in D(M)$ so that $\alpha b_{\leq k} \in D^*_{k+1}$.
- If $\varphi_k = 0$, then we may choose any $b_k \in D(M)$ that we desire. Since this happens infinitely often, we may ensure that $\mathrm{dcl}(b) = M$.

In the end, $\alpha b \in D^*_\Phi$ and $\mathrm{dcl}(\alpha b) = \mathrm{dcl}(b) = M$, so αb codes M. Taking n_0 large enough, αb is as close as desired to αa.

This argument shows, in particular, that there exists $\alpha a \in D(M)$ that codes M. Let $\alpha_n = \alpha/2^n$. Then $\alpha_n a \in D(M)$ codes M for each n, and $\alpha_n a \to 0$, so the root can also be approximated by codes for M. □

Definition 7.3. Let T be any theory in a countable language, and D^* its universal star sort. View it as a coding sort, as per Theorem 7.2, with exceptional set $D^0 = \{0\}$, and define the corresponding groupoid, as per Definition 3.1:

$$\boldsymbol{G}^*(T) = \boldsymbol{G}_{D^*}(T).$$

We already know that this is an open Polish topological groupoid, with basis $\boldsymbol{B}^*(T) \simeq \mathrm{S}_{D^*}(T)$.

Theorem 7.4. *The groupoid $\boldsymbol{G}^*(T)$ is a complete bi-interpretation invariant for the class of theories in countable languages.*

Proof. On the one hand, by Corollary 6.10, D^* only depends on the bi-interpretation class of T, and therefore so does $\boldsymbol{G}^*(T)$. Conversely, by Theorem 3.9, a theory bi-interpretable with T (namely, the theory T_{2D^*}, up to some arbitrary choices of definable distance and symbols for the language) can be recovered from $\boldsymbol{G}^*(T)$. □

Our last task is to calculate the basis $\mathrm{S}_{D^*}(T)$ explicitly, and show how Theorem 7.4 extends previous results, in a style similar to that of Remark 3.13.

Let us fix a rich sequence Φ on a sort D, so we may take $D^* = D^*_\Phi$. We also fix a formula $\chi(y)$ on D that is onto $[0, 1]$. Finally, we may assume that $\varphi_n(x_{<n}, y) = n\chi(y)$ for infinitely many n.

Let $X = \mathrm{S}_{D^{\mathbf{N}}}(T)$ and $Y = \mathrm{S}_{D^*_\Phi}(T)$. We may identify $\mathrm{S}_{*D^{\mathbf{N}}}(T)$ with $*X$, identifying $\mathrm{tp}(\alpha x)$ with $\alpha\,\mathrm{tp}(x)$ (here we need to assume that T is complete, so there exists a unique possible complete type for $0 \in D^*_\Phi$). This identifies Y with a subset of $*X$, namely that of all αp where $p(x)$ implies that $\alpha x \in D^*_\Phi$, or equivalently, such that $\alpha\varphi_n(p) \leq 1$ for all n.

For $\alpha \in [0, 1]$, let

$$X_\alpha = \{p \in X : \alpha p \in Y\}.$$

In particular, $X_0 = X$. Define $\rho : X \to [0, 1]$ by

$$\rho(p) = \sup\{\alpha : \alpha p \in Y\} = \sup\{\alpha : p \in X_\alpha\}.$$

Lemma 7.5. *Let $\alpha > 0$. Then for every $p \in X$ we have $\alpha \leq \rho(p)$ if and only if $p \in X_\alpha$, and X_α is compact, totally disconnected. In particular, $\rho : X \to [0, 1]$ is upper semicontinuous.*

Proof. For the first assertion, it is enough to notice that by compactness, the supremum is attained, namely, $p \in X_{\rho(p)}$. It follows that the condition $\rho(p) \geq \alpha$ is equivalent to $p \in X_\alpha$, so it is closed, and ρ is upper semicontinuous.

Assume that $\alpha q_i \in Y$ and $q_0 \neq q_1$. Then for some finite n, there exists a formula $\psi(x_{<n})$ that separates q_0 from q_1, say $\psi(q_i) = i$. We may also find a $[0, 1]$-valued formula $\chi(y)$ on D that attains (at least) the values 0 and 1.

By Urysohn's lemma, there exists a formula $\varphi(x_{<n}, y) \geq 0$ such that

$$|\psi(x_{<n}) + \chi(y) - 1| \geq \frac{1}{3} \qquad \Longrightarrow \qquad \varphi(x_{<n}, y) = 0,$$
$$|\psi(x_{<n}) + \chi(y) - 1| \leq \frac{1}{6} \qquad \Longrightarrow \qquad \varphi(x_{<n}, y) = \frac{17}{\alpha} + 42.$$

Since χ attains both 0 and 1, the formula $\varphi(x_{<n}, y)$ is witness-normalised, so there exists $k \geq n$ with $|\varphi - \varphi_k| \leq 1$.

Assume now that $\alpha p \in Y$. Then $\varphi_k(x_{<k}, x_k)^p \leq 1/\alpha$, so

$$\varphi(x_{<n}, x_k)^p \leq \frac{1}{\alpha} + 1 < \frac{17}{\alpha} + 42 \quad \text{and} \quad |\psi(x_{<n}) + \chi(x_k) - 1| > \frac{1}{6}.$$

This splits the set X_α in two (cl)open sets, defined by $\psi(x_{<n}) + \chi(x_k) > \frac{7}{6}$ and $\psi(x_{<n}) + \chi(x_k) < \frac{5}{6}$, respectively. Since χ is $[0, 1]$-valued, q_0 must belong to the latter and q_1 to the former, so they can be separated in X_α by clopen sets, completing the proof. □

Lemma 7.6. *The set $X_{>0} = \{p \in X : \rho(p) > 0\} = \bigcup_{\alpha>0} X_\alpha$ is totally disconnected, admitting a countable family of clopen sets $(U_n : n \in \mathbf{N})$ that separates points.*

Proof. We may write $X_{>0}$ as $\bigcup_k X_{2^{-k}}$. Each $X_{2^{-k}}$ is compact, totally disconnected, and it is metrisable by countability of the language. Therefore, it admits a basis of clopen sets.

The inclusion $X_{2^{-k}} \subseteq X_{2^{-k-1}}$ is a topological embedding of compact totally disconnected spaces. Therefore, if $U \subseteq X_{2^{-k}}$ is clopen, then we may find a clopen $U' \subseteq X_{2^{-k-1}}$ such that $U' \cap X_{2^{-k}} = U$. Proceeding in this fashion, we may find a clopen $\overline{U} \subseteq X_{>0}$ such that $\overline{U} \cap X_{2^{-k}} = U$.

We can therefore produce a countable family of clopen sets $(U_n : n \in \mathbf{N})$ in $X_{>0}$ such that for each k, $(U_n \cap X_{2^{-k}} : n \in \mathbf{N})$ is a basis of clopen sets for $X_{2^{-k}}$, and in particular separates points. It follows that (U_n) separates points in $X_{>0}$. □

Given this family (U_n), we may define a map $\theta_0 : X_{>0} \to 2^{\mathbf{N}}$, where $\theta_0(p)_n = 0$ if $p \in U_n$ and $\theta_0(p)_n = 1$ otherwise. It is continuous by definition, and injective since the sequence (U_n) separates points. If $\alpha p \in Y$, then either $\alpha = 0$ or $p \in X_{>0}$ (or possibly both), and we may define

$$\theta(\alpha p) = \alpha\theta_0(p) \in *2^{\mathbf{N}},$$

where $\theta(0) = \theta(0 \cdot p) = 0$ regardless of p. It is clearly continuous at 0, and at every point of Y (since θ_0 is continuous). It is also injective on Y. Since Y is compact, $\theta : Y \to *2^{\mathbf{N}}$ is a topological embedding.

Lemma 7.7. *The set of $\rho(p)p$ for $p \in X_{>0}$ is dense in Y.*

Proof. We already know that $\rho(p)p \in Y$. Assume now that $U \subseteq Y$ is open and nonempty, so it must contain some point αp with $\alpha > 0$.

We may assume that

$$U = \{\beta q \in Y : |\beta - \alpha| < \varepsilon,\ q \in V\},$$

where V is an open neighbourhood of p in X. The set V may be taken to be defined by a condition $\psi > 0$, where $\psi(x_{<n})$ only involves finitely many variables. By hypothesis on Φ, possibly increasing n, we may assume that $\varphi_n(x_{<n}, y) = n\chi(y)$, and we may further assume that $\alpha > 1/n$.

Choose a realisation a of p. Let $b_{<n} = a_{<n}$ and choose b_n so $\chi(b_n) = 1/n\alpha$. Then $\varphi_n(b_{<n}, b_n) = 1/\alpha$, so $\rho_{n+1}(b_{\le n}) = \alpha$, and we may extend $b_{\le n}$ to a sequence b such that $\rho(x') = \alpha$. In particular, $q = \mathrm{tp}(b) \in V \cap X_{>0}$ and $\alpha q = \rho(q)q \in U$. □

Let us recall from [Charatonik 1989] a few definitions and facts regarding fans. The *Cantor fan* is the space $*2^{\mathbf{N}}$. It is a connected compact metrisable topological space. More generally, a *fan* F is a connected compact space that embeds in the Cantor fan. An *endpoint* of F is a point $x \in F$ such that $F \smallsetminus \{x\}$ is connected (or empty, in the extremely degenerate case where F is reduced to a single point). If the set of endpoints is dense in F, then F is a *Lelek fan*. By the main theorem of [Charatonik 1989], the Lelek fan is unique up to homeomorphism.

Proposition 7.8. *Let T be a complete theory. Then $\mathrm{S}_{D^*}(T)$, the type-space of the universal star sort D^* in T, is homeomorphic to the Lelek fan.*

Proof. By Lemmas 7.5–7.7, the space $\mathrm{S}_{D^*}(T)$ is a Lelek fan. □

This gives us a hint as to how to relate the universal star sort with previously known coding sorts referred to in the examples of Section 3.

Theorem 7.9. *Assume T admits a universal Skolem sort D in the sense of* [Ben Yaacov 2022], *and let L denote the Lelek fan. Then $L * D$ is a universal star sort.*

Proof. We may assume that $L \subseteq *2^{\mathbf{N}}$, and moreover, that for every nonempty open subset $U \subseteq 2^{\mathbf{N}}$ there exist $\alpha > 0$ and $t \in U$ such that $\alpha t \in L$ (otherwise, we may replace $2^{\mathbf{N}}$ with the intersection of all clopen subsets for which this is true).

For each $n \in \mathbf{N}$ there is a natural initial projection $2^{\mathbf{N}} \to 2^n$. This induces in turn a projection $*2^{\mathbf{N}} \to *2^n$. Let $L_n \subseteq *2^n$ be the image of L under this projection, so $L = \varprojlim L_n$. Consequently, $L * D = \varprojlim (L_n * D)$.

Our hypotheses regarding L implies that the endpoints of L_n can be enumerated as $\{\alpha_t t : t \in 2^n\}$, with $\alpha_t > 0$. If $m \geq n$, then we have a natural projection $L_m \to L_n$. If $t \in 2^n$, $s \in 2^{m-n}$, and $ts \in 2^m$ is the concatenation, then $\alpha_{ts} ts$ gets sent to $\alpha_{ts} t \in L_n$, so $\alpha_{ts} \leq \alpha_t$, and $\alpha_{ts} = \alpha_t$ for at least one s. For any $\delta > 0$, we may always choose m large enough such that for every $t \in 2^n$, the set $\{\alpha_{ts} : s \in 2^{m-n}\}$ is δ-dense in the interval $[0, \alpha_t]$.

Let $\varphi(u, v)$ be a star correspondence between $L_n * D$ and some other star sort E^*, and let $\varepsilon > 0$. Choose $\delta > 0$ appropriately, and a corresponding m as in the previous paragraph. Define a formula on $2^n \times 2^{m-n} \times D \times E^*$ by

$$\varphi'(ts, x, v) = \varphi(\alpha_{ts} t * x, v).$$

On the one hand, since φ is witness-normalised in the first argument, φ' is witness-normalised in (ts, x). On the other hand, if $v \in E^*$, then there exist $\alpha t \in L_n$ (so $\alpha \leq \alpha_t$) and $x \in D$ (possibly in an elementary extension) such that $\varphi(\alpha t * x, v) = 0$. Having chosen δ small enough to begin with, and m large enough accordingly, we may now find $s \in 2^{m-n}$ such that α_{ts} is close to α, sufficiently so that

$$\varphi'(ts, x, v) = \varphi(\alpha_{ts} t * x, v) < \varepsilon.$$

It follows that $\varphi' \dot{-} \varepsilon$ is witness-normalised in either (ts, x) or v.

Let us now evoke a few black boxes from [Ben Yaacov 2022]. First, $2^m \times D$ is again a universal Skolem sort (and therefore stands in definable bijection with D). Second, since $\varphi' \dot{-} \varepsilon$ is witness-normalised in either group of arguments, there exists a surjective definable function $\sigma : 2^m \times D \to E^*$ that satisfies

$$(\varphi' \dot{-} \varepsilon)(ts, x, \sigma(ts, x)) \leq \varepsilon,$$

i.e., $\varphi'(ts, x, \sigma(ts, x)) \leq 2\varepsilon$. Define on $L_m * D \times E^*$

$$\psi(\alpha ts * x, v) = d(v, \alpha\alpha_{ts}^{-1} \sigma(ts, x))$$

(keeping in mind that if $\alpha ts \in L_m$, then $\alpha \leq \alpha_{ts}$). This formula is jointly sub-homogeneous (since d is, on E^*). It is also witness-normalised in $\alpha ts * x$ (just choose $v = \alpha\alpha_{ts}^{-1}\sigma(ts, x)$), and in v (since σ is surjective, and we may always choose $\alpha = \alpha_{ts}$). By construction, $\varphi(\alpha_{ts} t * x, \sigma(ts, x)) \leq 2\varepsilon$, so multiplying all arguments by $\alpha\alpha_{ts}^{-1}$,

$$\varphi(\alpha t * x, \alpha\alpha_{ts}^{-1}\sigma(ts, x)) \leq 2\varepsilon.$$

Therefore, if $\psi(\alpha ts * x, v)$ is small enough, $\varphi(\alpha t * x, v) \leq 3\varepsilon$, and by definition, if $\psi(\alpha ts * x, v_i)$ is small for $i = 0, 1$, then $d(v_0, v_1)$ is small. Replacing ψ with a multiple, we may replace "small enough" with "smaller than one", and now, by Lemma 5.11, $L * D$ is a universal star sort. □

Corollary 7.10. *Assume that T is $\aleph_0$-categorical and let D_0 be as in Example 3.10. In other words, let $M \vDash T$ be the separable model, $a \in M^{\mathbf{N}}$ a dense sequence, and D_0 the collection of realisations of* $\mathrm{tp}(a)$. *Then D_0 is a definable set, i.e., a sort, and $L * D_0$ is a universal star sort.*

Proof. In an $\aleph_0$-categorical theory, every type-definable set is definable. By [Ben Yaacov 2022, Proposition 4.17], $2^{\mathbf{N}} \times D_0$ is a universal Skolem sort. Now, $L * 2^{\mathbf{N}} \subseteq (*2^{\mathbf{N}}) * 2^{\mathbf{N}} = *(2^{\mathbf{N}} \times 2^{\mathbf{N}})$ is easily checked to be a fan, whose set of endpoints is dense, so it is homeomorphic to L. Therefore

$$L * (2^{\mathbf{N}} \times D_0) = (L * 2^{\mathbf{N}}) * D_0 \simeq L * D_0.$$

By Theorem 7.9, this is a universal star sort. □

Define $L^{(2)} \subseteq L^2$ as the set of pairs (x, y) such that either both $x = y = 0$, or both are nonzero. This is a Polish, albeit noncompact, star space, with root $(0, 0)$. When $\boldsymbol{G}$ is a topological groupoid, we may equip $L^{(2)} * \boldsymbol{G}$ with a groupoid composition law

$$[x, y, g] \cdot [y, z, h] = [x, z, gh].$$

If $\boldsymbol{B}$ is the basis of $\boldsymbol{G}$, then $L * \boldsymbol{B}$ is the basis of $L^{(2)} * \boldsymbol{G}$.

Corollary 7.11. *Let T be a continuous theory admitting a universal Skolem sort D, and let $\boldsymbol{G}(T) = \boldsymbol{G}_D(T)$, as in Example 3.12. Then $\boldsymbol{G}^*(T) \simeq L^{(2)} * \boldsymbol{G}(T)$. If T is $\aleph_0$-categorical, and $G(T)$ is the automorphism group of its unique separable model, then $\boldsymbol{G}^*(T) \simeq L^{(2)} * G(T)$.*

Proof. Just put the identities $D^* = L * D$ and $D^* = L * D_0$ through the groupoid construction. □

References

[Ahlbrandt and Ziegler 1986] G. Ahlbrandt and M. Ziegler, "Quasi-finitely axiomatizable totally categorical theories", *Ann. Pure Appl. Logic* **30**:1 (1986), 63–82. MR Zbl

[Ben Yaacov 2010] I. Ben Yaacov, "Definability of groups in $\aleph_0$-stable metric structures", *J. Symbolic Logic* **75**:3 (2010), 817–840. MR Zbl

[Ben Yaacov 2022] I. Ben Yaacov, "Reconstruction of non-$\aleph_0$-categorical theories", *J. Symb. Log.* **87**:1 (2022), 159–187. MR Zbl

[Ben Yaacov and Kaïchouh 2016] I. Ben Yaacov and A. Kaïchouh, "Reconstruction of separably categorical metric structures", *J. Symb. Log.* **81**:1 (2016), 216–224. MR Zbl

[Ben Yaacov and Usvyatsov 2010] I. Ben Yaacov and A. Usvyatsov, "Continuous first order logic and local stability", *Trans. Amer. Math. Soc.* **362**:10 (2010), 5213–5259. MR Zbl

[Ben Yaacov et al. 2008] 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

[Charatonik 1989] W. J. Charatonik, "The Lelek fan is unique", *Houston J. Math.* **15**:1 (1989), 27–34. MR Zbl

[Hanson 2020] J. E. Hanson, *Definability and categoricity in continuous logic*, Ph.D. thesis, University of Wisconsin–Madison, 2020, available at https://www.proquest.com/docview/2476546123. MR

[Mackenzie 1987] K. Mackenzie, *Lie groupoids and Lie algebroids in differential geometry*, London Math. Soc. Lecture Note Ser. **124**, Cambridge Univ. Press, 1987. MR Zbl

Received 25 Mar 2022. Revised 29 Sep 2022.

ITAÏ BEN YAACOV: Institut Camille Jordan, CNRS UMR 5208, Université Claude Bernard – Lyon 1, Villeurbanne, France

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.325

An improved bound for regular decompositions of 3-uniform hypergraphs of bounded VC_2-dimension

Caroline Terry

A regular partition $\mathcal{P}$ for a 3-uniform hypergraph $H = (V, E)$ consists of a partition $V = V_1 \cup \cdots \cup V_t$ and for each $ij \in \binom{[t]}{2}$, a partition $K_2[V_i, V_j] = P_{ij}^1 \cup \cdots \cup P_{ij}^\ell$ such that certain quasirandomness properties hold. The *complexity of* $\mathcal{P}$ is the pair (t, ℓ). In this paper we show that if a 3-uniform hypergraph H has VC_2-dimension at most k, then there is such a regular partition $\mathcal{P}$ for H of complexity (t, ℓ), where ℓ is bounded by a polynomial in the degree of regularity. This is a vast improvement on the bound arising from the proof of this regularity lemma in general, in which the bound generated for ℓ is of Wowzer type. This can be seen as a higher arity analogue of the efficient regularity lemmas for graphs and hypergraphs of bounded VC-dimension due to Alon–Fischer–Newman, Lovász–Szegedy, and Fox–Pach–Suk.

1. Introduction

Szemerédi's regularity lemma is an important theorem with many applications in extremal combinatorics. The proof of the regularity lemma, which first appeared in the 70s [24], was well known to produce tower-type bounds in ϵ. The question of whether this type of bound is necessary was resolved in the late 90s by Gowers' lower bound construction [10], which showed tower bounds are indeed required (see also [7; 17; 6]).

Hypergraph regularity was developed in the 2000s by Frankl, Gowers, Kohayakawa, Nagle, Rödl, Skokan, Schacht [9; 11; 12; 21; 20; 19], in order to prove a general counting lemma for hypergraphs. These types of regularity lemmas are substantially more complicated than prior regularity lemmas. In particular, a regular partition of a k-uniform hypergraph involves a sequence $\mathcal{P}_1, \ldots, \mathcal{P}_{k-1}$, where $\mathcal{P}_i$ is a collection of subsets $\binom{V}{i}$ such that certain quasirandomness properties hold for each $\mathcal{P}_i$ relative to $\mathcal{P}_1, \ldots, \mathcal{P}_{i-1}$. The proofs of these strong regularity lemmas produce Ackerman style bounds for the size of each $\mathcal{P}_i$. Given a function f, let $f^{(i)}$ denote the i-times iterate of f. We then define $\mathrm{Ack}_1(x) = 2^x$, and for $i > 1$,

The author was partially supported by NSF grant DMS-2115518.
MSC2020: 03C45, 05D99.
Keywords: model theory, VC-dimension, regularity lemmas, hypergraph regularity.

$\mathrm{Ack}_i(x) = \mathrm{Ack}_{k-1}^{(i)}(x)$. The proofs of the strong regularity lemma for k-uniform hypergraphs produce bounds for the size of each $\mathcal{P}_i$ of the form Ack_k. It was shown by Moshkovitz and Shapira [18] that this type of bound is indeed necessary for the size of $\mathcal{P}_1$, which corresponds to the partition of the vertex set.

In the case of 3-uniform hypergraphs, a decomposition in this sense consists of a partition $\mathcal{P}_1 = \{V_1, \dots, V_t\}$ of V, and a set

$$\mathcal{P}_2 = \left\{P_{ij}^{\alpha} : ij \in \tbinom{[t]}{2}, \alpha \in [\ell]\right\},$$

where for each $ij \in \binom{[t]}{2}$, $P_{ij}^1 \cup \dots \cup P_{ij}^{\ell}$ is a partition of $K_2[V_i, V_j]$. The *complexity* of $\mathcal{P}$ is the pair (t, ℓ). We give a formal statement of the regularity lemma for 3-graphs here for reference, and refer the reader to Section 2B for the precise definitions involved. The version stated below is a refinement of a regularity lemma due to Gowers [12] (for more details see Section 2B).

Theorem 1.1 (strong regularity lemma for 3-graphs). *For all $\epsilon_1 > 0$, and every function $\epsilon_2 : \mathbb{N} \to (0, 1]$, there exist positive integers T_0, L_0, and n_0 such that for any 3-graph $H = (V, E)$ on $n \geq n_0$ vertices, there exists a $\mathrm{dev}_{2,3}(\epsilon_1, \epsilon_2(\ell))$-regular, $(t, \ell, \epsilon_1, \epsilon_2(\ell))$-decomposition $\mathcal{P}$ for H with $t \leq T_0$ and $\ell \leq L_0$.*

In Theorem 1.1, the parameter T_0 is the bound for t, the size of the vertex partition, and L_0 is the bound for ℓ, the size of the partition of $K_2[V_i, V_j]$, for each $ij \in \binom{[t]}{2}$. The proof of Theorem 1.1 generates a Wowzer (i.e., Ack_3) type bound for both t and ℓ. Moshkovitz and Shapira showed in [18] that there exist 3-uniform hypergraphs requiring a Wowzer type bound for the size of t in Theorem 1.1. Less attention has been paid to the form of the bound L_0, and it remains open whether this is necessarily of Wowzer type. In recent work of the author and Wolf [26], the partition $\mathcal{P}_2$ plays a crucial role in the proof of a strong version of Theorem 1.1 in a combinatorially tame setting. This work suggests that understanding the form of the bound for ℓ is also an interesting problem.

In the case of graphs, it was shown that dramatic improvements on the bounds in Szemerédi's regularity lemma can be obtained under the hypothesis of bounded VC-dimension. In particular, Alon, Fischer, and Newman [1] showed that if a bipartite graph G has VC-dimension less than k, the it has an ϵ-regular partition of size at most $(k/\epsilon)^{O(k)}$. Lovász and Szegedy [16] extended this to all graphs of VC-dimension less than k, with a bound of the form $\epsilon^{-O(k^2)}$. Fox, Pach, and Suck [8] strengthened the bound to one of the form $c(\epsilon^{2k-1})$, and extended these results to hypergraphs of bounded VC-dimension. Related results were obtained with weaker polynomial bounds by Chernikov and Starchenko [3].

In this paper we prove an analogous theorem in the context of strong regularity for 3-uniform hypergraphs, where VC-dimension is replaced by a higher arity analogue called VC_2-dimension.

Definition 1.2. Suppose $H = (V, E)$ is a 3-graph. The VC$_2$*-dimension of* H, $\mathrm{VC}_2(H)$, is the largest integer k so that there exist $a_1, \dots, a_k, b_1, \dots, b_k \in V$ and $c_S \in V$ for each $S \subseteq [k]^2$, such that $a_i b_j c_S \in E$ if and only if $(i, j) \in S$.

The notion of VC$_2$-dimension was first introduced by Shelah [22], who also studied it in the context of groups [23]. It was later shown to have nice model-theoretic characterizations by Chernikov, Palacin, and Takeuchi [5] to have further natural connections to groups and fields by Hempel and Chernikov [15; 2], and to have applications in combinatorics by the author [25].

Using infinitary techniques, Chernikov and Towsner [4] proved a strong regularity lemma for 3-uniform hypergraphs of bounded VC$_2$-dimension without explicit bounds (in fact they proved results for k-uniform hypergraphs of bounded VC$_{k-1}$-dimension). Similar results were proved by the author and Wolf [26] in the 3-uniform case with Wowzer type bounds. In this paper, we show that 3-uniform hypergraphs of uniformly bounded VC$_2$-dimension have regular decompositions with vastly improved bounds on the size of ℓ; in particular, ℓ can be guaranteed to be polynomial in size, rather than Wowzer. We include the formal statement of our main theorem below, and refer the reader to the next section for details on the definitions involved.

Theorem 1.3. *For all* $k \geq 1$, *there are* $\epsilon_1^* > 0$ *and* $\epsilon_2^* : \mathbb{N} \to (0, 1]$ *such that the following holds. Suppose* $0 < \epsilon_1 < \epsilon_1^*$ *and* $\epsilon_2 : \mathbb{N} \to (0, 1]$ *satisfies* $0 < \epsilon_2(x) < \epsilon_2^*(x)$ *for all* $x \in \mathbb{N}$. *There is* $T = T(\epsilon_1, \epsilon_2)$ *such that every sufficiently large* 3*-graph* $H = (V, E)$ *has a* $\mathrm{dev}_{2,3}(\epsilon_1, \epsilon_2(\ell))$*-regular* $(t, \ell, \epsilon_1, \epsilon_2(\ell))$*-decomposition with* $\ell \leq \epsilon_1^{-O_k(k)}$ *and* $t \leq T$.

The bound T in Theorem 1.3 is generated from an application of Theorem 1.1, and is also of Wowzer type (see Theorem 3.1 for a more precise statement regarding this). The regular partition in Theorem 1.3 has the additional property that the regular triads have edge densities near 0 or 1, which also occurs in the results from [4; 26]. The ingredients in the proof of Theorem 1.3 include the improved regularity lemma for 3-graphs of bounded VC$_2$-dimension from [26], a method of producing quotient graphs from regular partitions of 3-graphs developed in [26], and ideas from [8] for producing weak regular partitions of hypergraphs of bounded VC-dimension.

The fact that the bound for ℓ can be brought all the way down to polynomial in Theorem 1.3 is somewhat surprising, given that the proof for arbitrary hypergraphs yields a Wowzer bound. This raises the question of what the correct form of the bound is, in general, for ℓ. The author conjectures it is at least a tower function (i.e., Ack$_2$).

It was conjectured in [4] that the bound for t can also be made sub-Wowzer under the assumption of bounded VC$_2$-dimension, however, the author has been unable to prove this is the case. This leaves the following open problem.

Problem 1.4. Given a fixed integer $k \geq 1$, are there arbitrarily large 3-uniform hypergraphs of VC_2-dimension at most k which require Wowzer type bounds for T_0 in Theorem 1.1?

2. Preliminaries

In this section we cover the requisite preliminaries, including graph and hypergraph regularity (Section 2B), VC and VC_2-dimension (Sections 2C, 2E, and 2F), auxiliary graphs defined from regular decompositions of 3-graphs (Section 2D), and basic lemmas around regularity and counting (Section 2G).

2A. ***Notation.*** We include here some basic notation needed for the other preliminary sections. Given a set V and $k \geq 1$, let

$$\binom{V}{k} = \{X \subseteq V : |X| = k\}.$$

A *k-uniform hypergraph* is a pair (V, E) where $E \subseteq \binom{V}{k}$. For a k-uniform hypergraph G, $V(G)$ denotes the vertex set of V and $E(G)$ denotes the edge set of G. Throughout the paper, all vertex sets are assumed to be finite.

When $k = 2$, we refer to a k-uniform hypergraph as simply a *graph*. When $k = 3$, we refer to a k-uniform hypergraph as a 3-*graph*.

Given distinct elements x, y, we write xy for the set $\{x, y\}$. Similarly, for distinct x, y, z, we write xyz for the set $\{x, y, z\}$. Given sets X, Y, Z, we set

$$K_2[X, Y] = \{xy : x \in X, y \in Y, x \neq y\} \quad \text{and}$$
$$K_3[X, Y, Z] = \{xyz : x \in X, y \in Y, z \in Z, x \neq y, y \neq z, x \neq z\}.$$

If $G = (V, E)$ is a graph and $X, Y \subseteq V$ are disjoint, we let $G[X, Y]$ be the bipartite graph $(X \cup Y, E \cap K_2[X, Y])$.

Given a k-uniform hypergraph $G = (V, E)$, $1 \leq i < k$, and $e \in \binom{V}{i}$, set

$$N_E(e) = \left\{e' \in \binom{V}{k-i} : e \cup e' \in E\right\}.$$

A *bipartite edge-colored graph* is a tuple $G = (A \cup B, E_0, E_1, \dots, E_i)$, where $i > 1$ and $K_2[A, B] = E_0 \sqcup E_1 \sqcup \cdots \sqcup E_i$. In this case, given $u \in \{0, 1, \dots, i\}$ and $x \in A \cup B$, we let $N_{E_u}(x) = \{y \in A \cup B : ab \in E_u\}$. Similarly, a *tripartite edge-colored* 3-*graph* is a tuple $G = (A \cup B \cup C, E_0, E_1, \dots, E_i)$, where $i > 1$ and $K_3[A, B, C] = E_0 \sqcup E_1 \sqcup \cdots \sqcup E_i$. In this case, given $u \in \{0, 1, \dots, i\}$ and $x, y \in V := A \cup B \cup C$, we let $N_{E_u}(x) = \{uv \in \binom{V}{2} : xuv \in E_u\}$ and $N_{E_u}(xy) = \{v \in V : xuv \in E_u\}$.

For two functions $f_1, f_2 : \mathbb{N} \to (0, 1]$, we write $f_1 < f_2$ to denote that $f_1(x) < f_2(x)$ for all $x \in \mathbb{N}$. For real numbers r_1, r_2 and $\epsilon > 0$, we write $r_1 = r_2 \pm \epsilon$ to denote that $r_1 \in (r_2 - \epsilon, r_2 + \epsilon)$. Given a natural number $n \geq 1$, $[n] = \{1, \dots, n\}$. An equipartition of a set V is a partition $V = V_1 \cup \cdots \cup V_t$ such that for each $1 \leq i, j \leq t$, we have $\big||V_i| - |V_j|\big| \leq 1$.

2B. *Regularity.* In this section we define graph regularity, as well as a strong notion of regularity for 3-graphs. We state our definitions in terms of the quasirandomness notion known as "dev", which is one of three notions of quasirandomness which are now known to be equivalent, the other two being "oct" and "disc". For more details on these and the equivalences, we refer the reader to [19].

We begin a notion of quasirandomness for graphs.

Definition 2.1. Suppose $B = (U \cup W, E)$ is a bipartite graph, and $|E| = d_B |U||W|$. We say B *has* $\mathrm{dev}_2(\epsilon, d)$ if $d_B = d \pm \epsilon$ and

$$\sum_{u_0,u_1 \in U} \sum_{w_0,w_1 \in W} \prod_{i \in \{0,1\}} \prod_{j \in \{0,1\}} g(u_i, v_j) \leq \epsilon |U|^2 |V|^2,$$

where $g(u, v) = 1 - d_B$ if $uv \in E$ and $g(u, v) = -d_B$ if $uv \notin E$.

We now define a generalization of Definition 2.1 to 3-graphs due to Gowers [11]. If $G = (V, E)$ is a graph, let $K_3^{(2)}(G)$ denote the set of triples from V forming a triangle in G, i.e.,

$$K_3^{(2)}(G) := \Big\{ xyz \in \binom{V}{3} : xy, yz, xz \in E \Big\}.$$

Now given a 3-graph $H = (V, R)$ on the same vertex set, we say that G *underlies* H if $R \subseteq K_3^{(2)}(G)$.

Definition 2.2. Assume $\epsilon_1, \epsilon_2 > 0$, $H = (V, E)$ is a 3-graph, $G = (U \cup W \cup Z, E)$ is a 3-partite graph underlying H, and $|E| = d_3 |K_3^{(2)}(G)|$. We say that (H, G) *has* $\mathrm{dev}_{2,3}(\epsilon_1, \epsilon_2)$ if there is $d_2 \in (0, 1)$ such that $G[U, W]$, $G[U, Z]$, and $G[W, Z]$ each have $\mathrm{dev}_2(\epsilon_2, d_2)$, and

$$\sum_{u_0,u_1 \in U} \sum_{w_0,w_1 \in W} \sum_{z_0,z_1 \in Z} \prod_{(i,j,k) \in \{0,1\}^3} h_{H,G}(u_i, w_j, z_k) \leq \epsilon_1 d_2^{12} |U|^2 |W|^2 |Z|^2,$$

where

$$h_{H,G}(x, y, z) = \begin{cases} 1 - d_3 & \text{if } xyz \in E \cap K_3^{(2)}(G), \\ -d_3 & \text{if } xyz \in K_3^{(2)}(G) \setminus E, \\ 0 & \text{if } xyz \notin K_3^{(2)}(G). \end{cases}$$

For the reader unfamiliar with hypergraph regularity, we note that in the notation of Definition 2.2, $d_2^{12} |U|^2 |W|^2 |Z|^2$ is approximately the number of tuples $(u_0, u_1, w_0, w_1, z_0, z_1) \in U^2 \times W^2 \times Z^2$ with $u_i w_j z_k \in K_3^{(2)}(G)$ for each $(i, j, k) \in \{0, 1\}^3$ (this is a consequence of the graph counting lemma and the assumption that $G[U,W]$, $G[U,Z]$, and $G[W,Z]$ have $\mathrm{dev}_2(\epsilon_2, d_2)$). Therefore, the first displayed equation in Definition 2.2 is bounding the quantity

$$\sum_{u_0,u_1\in U}\sum_{w_0,w_1\in W}\sum_{z_0,z_1\in Z}\prod_{(i,j,k)\in\{0,1\}^3} h_{H,G}(u_i,w_j,z_k)$$

in terms of total number of tuples $(u_0,u_1,w_0,w_1,z_0,z_1)\in U^2\times W^2\times Z^2$, where $\prod_{(i,j,k)\in\{0,1\}^3} h_{H,G}(u_i,w_j,z_k)$ is nonzero.

We now define a (t,ℓ)-decomposition for a vertex set V, which partitions V, as well as pairs from V.

Definition 2.3. Let V be a vertex set and $t,\ell\in\mathbb{N}^{>0}$. A (t,ℓ)-*decomposition* $\mathcal{P}$ for V consists of a partition $\mathcal{P}_1=\{V_1\cup\cdots\cup V_t\}$ of V, and for each $1\le i\ne j\le t$, a partition $K_2[V_i,V_j]=P^1_{ij}\cup\cdots\cup P^\ell_{ij}$. We let $\mathcal{P}_2=\{P^\alpha_{ij}: ij\in\binom{[t]}{2},\alpha\le\ell\}$.

A *triad of* $\mathcal{P}$ is a 3-partite graph of the form

$$G^{ijk}_{\alpha,\beta,\gamma}:=(V_i\cup V_j\cup V_k, P^\alpha_{ij}\cup P^\beta_{ik}\cup P^\gamma_{jk}),$$

for some $ijk\in\binom{[t]}{3}$ and $\alpha,\beta,\gamma\le\ell$. Let $\mathrm{Triads}(\mathcal{P})$ denote the set of all triads of $\mathcal{P}$, and observe that $\{K_3^{(2)}(G): G\in\mathrm{Triads}(\mathcal{P})\}$ partitions the set of triples $xyz\in\binom{V}{3}$ which are in distinct elements of $\mathcal{P}_1$.

For a 3-graph $H=(V,R)$, a decomposition $\mathcal{P}$ of V, and $G\in\mathrm{Triads}(\mathcal{P})$, define $H|G:=(V(G),R\cap K_3^{(2)}(G))$. Note that G always underlies $H|G$.

Definition 2.4. Given a 3-graph $H=(V,R)$, a decomposition $\mathcal{P}$ of V, and $G\in\mathrm{Triads}(\mathcal{P})$, we say G *has* $\mathrm{dev}_{2,3}(\epsilon_1,\epsilon_2)$ *with respect to* H if $(H|G,G)$ has $\mathrm{dev}_{2,3}(\epsilon_1,\epsilon_2)$.

To define a regular decomposition for a 3-graph, we need one more notion, namely that of an "equitable" decomposition.

Definition 2.5. We say that $\mathcal{P}$ is a $(t,\ell,\epsilon_1,\epsilon_2)$-*decomposition* if $\mathcal{P}_1=\{V_1,\dots,V_t\}$ is an equipartition and for at least $(1-\epsilon_1)\binom{|V|}{2}$ many $xy\in\binom{V}{2}$, there is some $P^\alpha_{ij}\in\mathcal{P}_2$ containing xy such that $(V_i\cup V_j,P^\alpha_{ij})$ has $\mathrm{dev}_2(\epsilon_2,1/\ell)$.

Definition 2.6. Suppose that $H=(V,E)$ is a 3-graph and $\mathcal{P}$ is a $(t,\ell,\epsilon_1,\epsilon_2)$-decomposition of V. We say that $\mathcal{P}$ is $\mathrm{dev}_{2,3}(\epsilon_1,\epsilon_2)$-*regular* for H if for all but at most $\epsilon_1 n^3$ many triples $xyz\in\binom{V}{3}$, the unique $G\in\mathrm{Triads}(\mathcal{P})$ with $xyz\in K_3^{(2)}(G)$ satisfies $\mathrm{dev}_{2,3}(\epsilon_1,\epsilon_2)$ with respect to H.

We can now restate the regularity lemma for $\mathrm{dev}_{2,3}$-quasirandomness.

Theorem 2.7. *For all* $\epsilon_1>0$, *every function* $\epsilon_2:\mathbb{N}\to(0,1]$, *and every* $\ell_0,t_0\ge 1$, *there exist positive integers* $T_0=T_0(\epsilon_1,\epsilon_2,t_0,\ell_0)$ *and* $L_0=L_0(\epsilon_1,\epsilon_2,t_0,\ell_0)$, *such that for every sufficiently large* 3*-graph* $H=(V,E)$, *there exists a* $\mathrm{dev}_{2,3}(\epsilon_1,\epsilon_2(\ell))$-*regular*, $(t,\ell,\epsilon_1,\epsilon_2(\ell))$-*decomposition* $\mathcal{P}$ *for* H *with* $t_0\le t\le T_0$ *and* $\ell_0\le\ell\le L_0$.

This theorem was first proved in a slightly different form by Gowers in [11]. In particular, in [11], the partition of the pairs $\mathcal{P}_2$ is not required to be equitable as it is in Theorem 2.7. Theorem 2.7 as stated appears in [19], where it is pointed out that the additional equitability requirement can be obtained using techniques from [9].

2C. VC-***dimension.*** In this subsection we give some preliminaries around VC and VC$_2$-dimension. We begin by defining VC-dimension.

Given a set V, $\mathcal{F} \subseteq \mathcal{P}(V)$, and $X \subseteq V$, let $|\mathcal{F} \cap X| := \{F \cap X : F \in \mathcal{F}\}$. We say that X is *shattered by* $\mathcal{F}$ if $|\mathcal{F} \cap X| = 2^{|X|}$. The *VC-dimension of* $\mathcal{F}$ is then defined to be the size of the largest subset of V which is shattered by $\mathcal{F}$.

For a graph $G = (V, E)$, the *VC-dimension of* G is the VC-dimension of the set system $\{N_E(x) : x \in V\} \subseteq \mathcal{P}(V)$. We now give a simple recharacterization of this. Given $k \geq 1$, let $A_k = \{a_i : i \in [k]\}$, and $C_{\mathcal{P}([k])} = \{c_S : S \subseteq [k]\}$.

Definition 2.8. For $k \geq 1$, define $U(k)$ to be the bipartite graph $(A_k \cup C_{\mathcal{P}([k])}, E)$, where $E = \{a_i c_S : i \in S\}$.

Then it is well known that a graph G has VC-dimension at least k if and only if there is a map $f : V(U(k)) \to V(G)$ so that for all $a \in A_k$ and $c \in C_{\mathcal{P}([k])}$, $ab \in E(U(k))$ if and only if $f(a)f(b) \in E(G)$.

2D. *Encodings.* In this subsection, we define an auxiliary edge-colored graph associated to a regular decomposition of a 3-graph. We then state a result from [26] which shows that encodings of $U(k)$ cannot occur when the auxiliary edge-colored graph arises from a regular decomposition of a 3-graph with VC$_2$-dimension less than k.

Definition 2.9. Suppose $\epsilon_1, \epsilon_2 > 0$, $\ell, t \geq 1$, V is a set, and $\mathcal{P}$ is a $(t, \ell, \epsilon_1, \epsilon_2)$-decomposition for V consisting of $\mathcal{P}_1 = \{V_i : i \in [t]\}$ and $\mathcal{P}_2 = \{P_{ij}^\alpha : ij \in \binom{[t]}{2}, \alpha \leq \ell\}$. Define

$$\mathcal{P}_{\mathrm{cnr}} = \Big\{P_{ij}^\alpha P_{ik}^\beta : ijk \in \binom{[t]}{3}, \alpha, \beta \leq \ell, \text{ and } P_{ij}^\alpha, P_{ij}^\beta \text{ satisfy } \mathrm{dev}_2(\epsilon_2, 1/\ell)\Big\},$$

$$\mathcal{P}_{\mathrm{edge}} = \{P_{ij}^\alpha \in \mathcal{P}_2 : P_{ij}^\alpha \text{ satisfies } \mathrm{dev}_2(\epsilon_2, 1/\ell)\}.$$

In the above, cnr stands for "corner". Observe that for each $P_{ij}^\alpha \in \mathcal{P}_{\mathrm{edge}}$ and $P_{uv}^\beta P_{uw}^\gamma \in \mathcal{P}_{\mathrm{cnr}}$, if $\{v, w\} = \{i, j\}$, then the pair $(P_{ij}^\alpha, P_{uv}^\beta P_{uw}^\gamma)$ corresponds to a triad from $\mathcal{P}$, namely G_{ijs}^{uvw}.

Definition 2.10. Suppose $\epsilon_1, \epsilon_2 > 0$, $\ell, t \geq 1$, $H = (V, E)$ is a 3-graph, and $\mathcal{P}$ is a $(t, \ell, \epsilon_1, \epsilon_2)$-decomposition for V. Define

$$\mathbf{E}_0 = \big\{P_{ij}^\alpha (P_{jk}^\beta P_{ik}^\gamma) \in K_2[\mathcal{P}_{\mathrm{edge}}, \mathcal{P}_{\mathrm{cnr}}] : |E \cap K_3^{(2)}(G_{ijk}^{\alpha\beta\gamma})| < \tfrac{1}{2}|K_3^{(2)}(G_{ijk}^{\alpha\beta\gamma})|\big\},$$

$$\mathbf{E}_1 = \big\{P_{ij}^\alpha (P_{jk}^\beta P_{ik}^\gamma) \in K_2[\mathcal{P}_{\mathrm{edge}}, \mathcal{P}_{\mathrm{cnr}}] : |E \cap K_3^{(2)}(G_{ijk}^{\alpha\beta\gamma})| \geq \tfrac{1}{2}|K_3^{(2)}(G_{ijk}^{\alpha\beta\gamma})|\big\}, \text{ and}$$

$$\mathbf{E}_2 = K_2[\mathcal{P}_{\mathrm{edge}}, \mathcal{P}_{\mathrm{cnr}}] \setminus (\mathbf{E}_1 \cup \mathbf{E}_0).$$

Note that Definition 2.10 gives us a natural bipartite edge-colored graph with vertex set $\mathcal{P}_{\text{edge}} \cup \mathcal{P}_{\text{cnr}}$ and edge sets given by $\mathbf{E}_0, \mathbf{E}_1, \mathbf{E}_2$. The author and Wolf showed in [26] that these auxiliary edge-colored graphs are useful for understanding 3-graphs of bounded VC_2-dimension. To explain why, we require the following notion of an "encoding".

Definition 2.11. Let $\epsilon_1, \epsilon_2 > 0$ and $t, \ell \geq 1$. Suppose $R = (A \cup B, E_R)$ is a bipartite graph, $H = (V, E)$ is a 3-graph, and $\mathcal{P}$ is a $(t, \ell, \epsilon_1, \epsilon_2)$-decomposition of V. An (A, B)-*encoding of R in* $(H, \mathcal{P})$ consists of a pair of functions (g, f), where $g : A \to \mathcal{P}_{\text{cnr}}$ and $f : B \to \mathcal{P}_{\text{edge}}$ are such that the following hold for some $j_0 k_0 \in \binom{[t]}{2}$:

(1) $\text{Im}(f) \subseteq \{P^{\alpha}_{j_0 k_0} : \alpha \leq \ell\}$, and $\text{Im}(g) \subseteq \{P^{\beta}_{i j_0} P^{\gamma}_{i k_0} : i \in [t], \beta, \gamma \leq \ell\}$.

(2) For all $a \in A$ and $b \in B$, if $ab \in E_R$, then $g(a) f(b) \in \mathbf{E}_1$, and if $ab \notin E_R$, then $g(a) f(b) \in \mathbf{E}_0$.

An *encoding of* $U(k)$ will always mean an $(A_k, C_{\mathcal{P}([k])})$-encoding of $U(k)$. In [26], we proved the following proposition connecting encodings of $U(k)$ and VC_2-dimension (see Theorem 6.5(2) in [26]).

Proposition 2.12. *For all $k \geq 1$, there are $\epsilon_1 > 0$ and $\epsilon_2 : \mathbb{N} \to (0, 1]$ such that for all $t, \ell \geq 1$, there is N such that the following hold. Suppose $H = (V, E)$ is a 3-graph with $|V| \geq N$, and $\mathcal{P}$ is a $\text{dev}_{2,3}(\epsilon_2(\ell), \epsilon_1)$-regular $(t, \ell, \epsilon_1, \epsilon_2(\ell))$-decomposition of V. If there exists an encoding of $U(k)$ in $(H, \mathcal{P})$, then H has k-IP_2.*

Moreover, there is a constant $C = C(k)$ so that $\epsilon_1 = (1/2)^C$.

We remark here that Proposition 2.12 is actually proved in [26] for an equivalent notion of quasirandomness called $\text{disc}_{2,3}$, and without the final "Moreover" statement regarding the quantitative form for ϵ_1 (see Proposition 5.6 in [26]). Tracing the bounds in the proof of Proposition 5.6 in [26], one finds that ϵ_1 has the form $\mu = \mu(\epsilon_1, k)$, where μ comes from a version of the counting lemma (see Theorem 3.1 in [26]). An explicit value for this μ is unclear, as the proof of the counting lemma for $\text{disc}_{2,3}$ passes through its equivalence with $\text{oct}_{2,3}$, and then the counting lemma for $\text{oct}_{2,3}$. The author has not found proofs of these results in the literature which are explicit in the parameters (see Corollary 2.3 in [19]). It seems that one could produce such an explicit result from [19] and [14] with some effort, however, we have instead chosen to side-step the issue by working with the quasirandomness notion dev, rather than disc.

In particular, all the ingredients used to prove Proposition 5.6 of [26] have well known analogues for dev. By running the same arguments as in [26] using dev rather than disc, one obtains Proposition 2.12 as stated. The additional "Moreover" statement about the explicit form for ϵ_1 then arises from the fact that there is a proof of the counting lemma for $\mathrm{dev}_{2,3}$ which is explicit in the parameters (see [11, Theorem 6.8]).

2E. *Haussler's packing lemma.* We will be applying techniques for proving improved regularity lemmas for graphs and hypergraphs of bounded VC-dimension to the edge-colored auxiliary graphs defined in the previous subsection. In particular, we will use ideas from the proof of Theorem 1.3 in [8]. We begin by describing the relevant result from VC-theory, namely Haussler's packing lemma.

Suppose V is a set and $\mathcal{F} \subseteq V$. We say that a subset $\mathcal{X} \subseteq \mathcal{F}$ is *δ-separated* if for all distinct $X, X' \in \mathcal{X}$, $|X \Delta X| > \delta$. The following packing lemma, due to Haussler, shows that if $\mathcal{F}$ has bounded VC-dimension, the size a of a δ-separated family cannot be too large [13].

Theorem 2.13 (Haussler's packing lemma). *Suppose $\mathcal{F} \subseteq \mathcal{P}(V)$, where $|V| = n$ and $\mathcal{F}$ has VC-dimension at most k. Then the maximal size of a δ-separated subcollection of $\mathcal{F}$ is at most $c_1(n/\delta)^k$, for some constant $c_1 = c_1(k)$.*

We will apply Theorem 2.13 in the setting of edge-colored graphs. This technique is inspired by the proof of Theorem 1.3 in [8].

Suppose $G = (A \cup B, E_0, E_1, E_2)$ is a bipartite edge-colored graph. We say that G has an E_0/E_1-copy of $U(k)$ if there are $v_1, \dots, v_k \in A$ and for each $S \subseteq [k]$ a vertex $w_S \in B$ such that $i \in S$ implies $v_i w_S \in E_1$ and $i \notin S$ implies $v_i w_S \in E_0$. Given $a, a' \in A$ and $\delta > 0$, write $a \sim_\delta a'$ if for each $u \in \{0, 1, 2\}$, $|N_{E_u}(a) \Delta N_{E_u}(a')| \le \delta |B|$. Our main application of Theorem 2.13 is the following lemma.

Lemma 2.14. *Suppose $k \ge 1$ and $c_1 = c_1(k)$ is as in Theorem 2.13. Suppose $d \ge 1$ and $\delta, \epsilon > 0$ satisfy $\epsilon \le c_1^{-2}(\delta/8)^{2k+2}$. Assume $G = (A \cup B, E_0, E_1, E_2)$ is a bipartite edge-colored graph, and assume there is no E_0/E_1-copy of $U(k)$ in G, and that $|E_2| \le \epsilon |A||B|$.*

Then there is an integer $m \le 2c_1(\delta/8)^{-k}$, vertices $x_1, \dots, x_m \in A$, and a set $U \subseteq A$ with $|U| \le \sqrt{\epsilon}\,|A|$, so that for all $a \in A \setminus U$, $|N_{E_2}(a)| \le \sqrt{\epsilon}\,|B|$ and there is some $1 \le i \le m$ so that $a \sim_\delta x_i$.

Proof. Let $U = \{v \in A : |N_{E_2}(v)| \ge \sqrt{\epsilon}\,|B|\}$. Since $|E_2| \le \epsilon |A||B|$, we know that $|U| \le \sqrt{\epsilon}\,|A|$. Let $A' = A \setminus U$. Let m be maximal such that there exist $x_1, \dots, x_m \in A'$, so that $\{N_{E_1}(x_i) : i \in [m]\}$ is a $\delta/2$-separated family of sets on B. We show $m \le 2c_1(\delta/8)^{-k}$.

Suppose towards a contradiction that $m \ge \lceil 2c_1(\delta/8)^{-k} \rceil$. Let

$$B' = B \setminus \Big(\bigcup_{i=1}^{m} E_2(x_i) \Big),$$

and let $\mathcal{F} := \{N_{E_1}(x_i) \cap B' : i \in [m]\}$. Notice $|B \setminus B'| \le m\sqrt{\epsilon}\,|B|$. We claim that $\mathcal{F}$ is $\delta/4$-separated. Consider $1 \le i \ne j \le m$. Then we know that

$$\begin{aligned}|N_{E_1}(x_i)\Delta N_{E_1}(x_j) \cap B'| &\ge |N_{E_1}(x_i)\Delta N_{E_1}(x_j)| - m\sqrt{\epsilon}\,|B| \\ &\ge |B|(\delta/2 - m\sqrt{\epsilon}) \\ &\ge |B|\delta/4,\end{aligned}$$

where the last inequality is by our assumptions on δ, ϵ. By Theorem 2.13, $\mathcal{F}$ shatters a set of size k. By construction, for each $1 \le i \le m$, $B' \setminus N_{E_1}(x_i) \subseteq N_{E_0}(x_i)$. Consequently, we must have that there exists an E_0/E_1-copy of $U(k)$ in G, a contradiction.

Thus, $m \le 2c_1(\delta/8)^{-k}$. For all $a \in A \setminus U$, we know that $|N_{E_2}(a)| \le \sqrt{\epsilon}\,|B|$, and there is some $1 \le i \le m$ so that $|N_{E_1}(a)\Delta N_{E_1}(x_i)| \le \delta|B|/2$. We claim that $a \sim_\delta x_i$. We already know that $|N_{E_1}(a) \cap N_{E_1}(x_i)| \le \delta|B|$. Since a, x_i are both in A', we have

$$|N_{E_2}(a)\Delta N_{E_2}(x_i)| \le |N_{E_2}(a)| + |N_{E_2}(a)| \le 2\sqrt{\epsilon}\,|B| < \delta|B|/2.$$

Combining these facts, we have that

$$|N_{E_0}(a)\Delta N_{E_0}(x_i)| \le |N_{E_2}(a)| + |N_{E_2}(a)| + |N_{E_1}(a)\Delta N_{E_1}(x_i)| \le \delta|B|.$$

Thus $a \sim_\delta x_i$, as desired. □

2F. *Tame regularity for 3-graphs of bounded* VC_2*-dimension.* In this subsection we state the tame regularity lemma for 3-graphs of bounded VC_2-dimension from [26].

Definition 2.15. Suppose $H = (V, E)$ is a 3-graph with $|V| = n$ and $\mu > 0$. Suppose $t, \ell \ge 1$ and $\mathcal{P}$ is a (t, ℓ)-decomposition of V. We say that $\mathcal{P}$ is *μ-homogeneous with respect to H* if at least $(1-\mu)\binom{n}{3}$ triples $xyz \in \binom{V}{3}$ satisfy the following: there is some $G \in \mathrm{Triads}(\mathcal{P})$ such that $xyz \in K_3^{(2)}(G)$ and either

$$|E \cap K_3^{(2)}(G)| \le \mu|K_3^{(2)}(G)| \quad \text{or} \quad |E \cap K_3^{(2)}(G)| \ge (1-\mu)|K_3^{(2)}(G)|.$$

Given a 3-graph $H = (V, E)$ and a $(t, \ell, \epsilon_1, \epsilon_2)$-decomposition $\mathcal{P}$ of V, we say that $\mathcal{P}$ is *μ-homogeneous with respect to H* if at least $(1-\mu)\binom{|V|}{3}$ triples $xyz \in \binom{V}{3}$ are in a μ-homogeneous triad of $\mathcal{P}$. We have the following theorem from [26].

Theorem 2.16. *For all $k \ge 1$, there are $\epsilon_1^* > 0$, $\epsilon_2^* : \mathbb{N} \to (0, 1]$, and a function $f : (0, 1] \to (0, 1]$ with $\lim_{x\to 0} f(x) = 0$ such that the following hold.*

Suppose $t_0, \ell_0 \ge 1$, $0 < \epsilon_1 < \epsilon_1^*$, and $\epsilon_2 : \mathbb{N} \to (0, 1]$ satisfies $\epsilon_2 < \epsilon_2^*$. Let N, T, and L be as in Theorem 2.7 for $\epsilon_1, \epsilon_2, t_0, \ell_0$. Suppose $H = (V, E)$ is a 3-graph with $|V| \ge N$ and $\mathrm{VC}_2(H) < k$. Then there exist $t_0 \le t \le T$, $\ell_0 \le \ell \le L$, *and a* $(t, \ell, \epsilon_1, \epsilon_2(\ell))$*-decomposition of* V *which is* $\mathrm{dev}_{2,3}(\epsilon_1, \epsilon_2(\ell))$*-regular and* $f(\epsilon_1)$*-homogeneous with respect to* H.

Moreover, f *may be taken to have the form* $x^{1/D}$, *where* $D \geq 1$ *depends only on* k.

Since the bounds in Theorem 2.16 come from Theorem 2.7, they are of Wowzer type. We also note that the proof of Theorem 2.16 in fact guarantees something slightly stronger, namely that every $\mathrm{dev}_{2,3}(\epsilon_1, \epsilon_2(\ell))$-regular triad of $\mathcal{P}$ is $f(\epsilon_1)$-homogeneous.

We remark here that Theorem 2.16 was proved in [26] for the notion of $\mathrm{disc}_{2,3}$ rather than $\mathrm{dev}_{2,3}$, and without the "moreover" statement regarding the form of the function f (see Proposition 3.2 in [26]). Examination of the proof of Proposition 3.2 in [26] shows that the function f depends on k and a version of the counting lemma for 3-graphs (namely Theorem 3.1 in [26]). An explicit expression for $f(x)$ in Proposition 3.2 of [26] would thus require a version of the counting lemma for $\mathrm{disc}_{2,3}$ which is explicit in the parameters. However, one can rerun all the arguments in [26] using the quasirandomness notion $\mathrm{dev}_{2,3}$ in place of $\mathrm{disc}_{2,3}$ to obtain Theorem 2.16 as stated. In this case, an explicit expression for f can be obtained using the counting lemma for $\mathrm{dev}_{2,3}$ (see also the discussion following Proposition 2.12).

2G. ***Other preliminaries.*** In this subsection we give several lemmas, most of which are basic facts about regularity and counting. First, we will use the following version of the triangle counting lemma.

Proposition 2.17 (counting lemma). *Suppose* $\epsilon, d > 0$. *Let* $G = (A \cup B \cup C, E)$ *be a* 3*-partite graph such that each of* $G[A, B]$, $G[B, C]$, *and* $G[A, C]$ *has* $\mathrm{dev}_2(\epsilon, d)$. *Then*

$$\big||K_3^{(2)}(G)| - d^3|A||B||C|\big| \leq 4\epsilon^{1/4}|A||B||C|.$$

For a proof, see [11, Lemma 3.4]. The following symmetry lemma was proved in [26] (see Lemma 4.9 there).

Lemma 2.18 (symmetry lemma). *For all* $0 < \epsilon < \frac{1}{4}$ *there is* n *such that the following holds. Suppose* $G = (U \cup W, E)$ *is a bipartite graph,* $|U|, |W| \geq n$, *and* $U' \subseteq U$, $W' \subseteq W$ *satisfy* $|U'| \geq (1-\epsilon)|U|$ *and* $|W'| \geq (1-\epsilon)|W|$. *Suppose that for all* $u \in U'$,

$$\max\{|N(u) \cap W|, |\neg N(u) \cap W|\} \geq (1-\epsilon)|W|,$$

and for all $w \in W'$,

$$\max\{|N(w) \cap U|, |\neg N(w) \cap U|\} \geq (1-\epsilon)|U|.$$

Then $|E|/|U||W| \in [0, 2\epsilon^{1/2}) \cup (1 - 2\epsilon^{1/2}, 1]$.

We will use the following immediate corollary of this.

Corollary 2.19. *For all $0 < \epsilon < \frac{1}{4}$ there is n such that the following holds. Suppose $G = (U \cup W, E)$ is a bipartite graph with $|U|, |W| \geq n$, and $|E|/|U||W| \in (2\epsilon^{1/2}, 1 - 2\epsilon^{1/2})$. Then one of the following hold.*

(1) *There is $U' \subseteq U$ with $|U'| \geq \epsilon|U|$, so that for all $u \in U$,*

$$\frac{|N_E(u) \cap W|}{|W|} \in (\epsilon, 1-\epsilon).$$

(2) *There is $W' \subseteq W$ with $|W'| \geq \epsilon|W|$, so that for all $w \in W$,*

$$\frac{|N_E(w) \cap U|}{|U|} \in (\epsilon, 1-\epsilon).$$

We will use a lemma which was originally proved by Frankl and Rödl (see [9, Lemma 3.8]) for another notion of quasirandomness for graphs, called disc_2.

Definition 2.20. Suppose $B = (U \cup W, E)$ is a bipartite graph, and $|E| = d_B|U||W|$. We say B *has* $\mathrm{disc}_2(\epsilon, d)$ if $d_B = d \pm \epsilon$ and for all $U' \subseteq U$ and $W' \subseteq W$,

$$\big||E \cap K_2[U', W']| - d|U'||W'|\big| \leq \epsilon|U||W|.$$

Gowers proved the following quantitative equivalence between disc_2 and dev_2 (see Theorem 3.1 in [11]).

Theorem 2.21. *Suppose $B = (U \cup W, E)$ is a bipartite graph. If B has* $\mathrm{disc}_2(\epsilon, d)$ *then it has* $\mathrm{dev}_2(\epsilon, d)$. *If B has* $\mathrm{dev}_2(\epsilon, d)$, *then it has* $\mathrm{disc}_2(\epsilon^{1/4}, d)$.

Combining Theorem 2.21 with Lemma 3.8 in [9], we obtain the following.

Lemma 2.22. *For all $\epsilon > 0$, $\rho \geq 2\epsilon$, $0 < p < \rho/2$, and $\delta > 0$, there is $m_0 = m_0(\epsilon, \rho, \delta)$ such that the following holds. Suppose $|U| = |V| = m \geq m_0$, and $G = (U \cup V, E)$ is a bipartite graph satisfying* $\mathrm{dev}_2(\epsilon)$ *with density ρ. Then if $\ell = [1/p]$ and $\epsilon \geq 10(1/\ell m)^{1/5}$, there is a partition $E = E_0 \cup E_1 \cup \cdots \cup E_\ell$ such that*

(1) *For each $1 \leq i \leq \ell$, $(U \cup V, E_i)$ has* $\mathrm{dev}_2(\epsilon^{1/4})$ *with density $\rho p(1 \pm \delta)$, and*

(2) $|E_0| \leq \rho p(1 + \delta)m^2$.

Further, if $1/p \in \mathbb{Z}$, then $E_0 = \varnothing$.

We will also use the following fact, which can be obtained from Fact 2.3 in [26] along with Theorem 2.21.

Fact 2.23. *Suppose E_1 and E_2 are disjoint subsets of $K_2[U, V]$. If $(U \cup V, E_1)$ has* $\mathrm{dev}_2(\epsilon_1, d_1)$, *and $(U \cup V, E_2)$ has* $\mathrm{dev}_2(\epsilon_2, d_2)$, *then $(U \cup V, E_1 \cup E_2)$ has* $\mathrm{dev}_2(\epsilon_1^{1/4} + \epsilon_2^{1/4}, d_2 + d_1)$.

Finally, we will use the fact that triads with density near 0 or 1 are quasirandom. For completeness, we include a proof of this in the Appendix.

Proposition 2.24. *For all $0 < \epsilon < \frac{1}{2}$, $d_2 > 0$, and $0 < \delta \leq (d_2/2)^{48}$, there is N such that the following holds. Suppose $H = (V_1 \cup V_2 \cup V_3, R)$ is a 3-partite 3-graph on $n \geq N$ vertices, and for each $i, j \in [3]$, $\big||V_i| - |V_j|\big| \leq \delta |V_i|$. Suppose $G = (V_1 \cup V_2 \cup V_3, E)$ is a 3-partite graph, where for each $1 \leq i < j \leq 3$, $G[V_i, V_j]$ has* $\mathrm{dev}_2(\delta, d_2)$, *and assume*

$$|R \cap K_3^{(2)}(G)| \leq \epsilon |K_3^{(2)}(G)|.$$

Then $(H|G, G)$ has $\mathrm{dev}_{2,3}(\delta, 6\epsilon)$.

3. Proof of main theorem

We first give a more precise statement of our main theorem.

Theorem 3.1. *For all $k \geq 1$, there are polynomials $p_1(x), p_2(x, y), p_3(x)$, a constant $\epsilon_1^* > 0$, and a function $\epsilon_2^* : \mathbb{N} \to (0, 1]$ such that the following holds, where $T_0(x, y, z, w)$ is as in Theorem 2.7.*

For all $0 < \epsilon_1 < \epsilon_1^*$ and $\epsilon_2 : \mathbb{N} \to (0, 1]$ satisfying $\epsilon_2 < \epsilon_2^*$, there is $L \leq \epsilon_1^{-O_k(k)}$ such that the following holds for $T = T_0(p_1(\epsilon_1), \epsilon_2 \circ q_2, p_3(\epsilon_1^{-1}), 1)$, where $q_2(y) = p_2(\epsilon_1, y)$.

Every sufficiently large 3-graph $H = (V, E)$ such that $\mathrm{VC}_2(H) < k$ has a $\mathrm{dev}_{2,3}(\epsilon_1, \epsilon_2(\ell))$-regular $(t, \ell, \epsilon_1, \epsilon_2(\ell))$-decomposition with $\ell \leq L$ and $t \leq T$.

We now give a few remarks regarding the bounds. As can be seen above, the bound T in Theorem 3.1 is obtained by composing the bound T_0 from Theorem 2.7 with several polynomial functions. This does not change the fundamental shape of the bound in terms of the Ackerman hierarchy, and thus the bound for t in Theorem 3.1 remains a Wowzer type function. On the other hand, we see that the bound for ℓ becomes polynomial in ϵ_1^{-1}.

The polynomial p_3 in Theorem 3.1 depends on the f in Theorem 2.16, which in turn depends on the hypergraph counting lemma for $\mathrm{dev}_{2,3}$. One could therefore obtain a quantitative version of Theorem 3.1 for the equivalent quasirandomness notions of $\mathrm{disc}_{2,3}$ and $\mathrm{oct}_{2,3}$ using the same arguments, given a quantitative version of their respective counting lemmas.

The general strategy for the proof of Theorem 3.1 is as follows. Given a large 3-graph H of VC$_2$-dimension less than k, we first apply Theorem 2.16 to obtain a homogeneous, regular partition $\mathcal{P}$ for H. We then consider the auxiliary edge-colored graphs associated to $\mathcal{P}$, as described in Section 2D. These contain no copies of $U(k)$ by Proposition 2.12, allowing us to apply Lemma 2.14. This yields decompositions for the auxiliary edge-colored graphs, which we eventually use to define a new decomposition $\mathcal{Q}$ for H which is still regular and homogeneous, but which has a polynomial bound for the parameter ℓ. This last part requires the most work, as well as most of the lemmas from Section 2G.

We have not sought to optimize constants which do not effect the overall form of the bounds involved.

Proof of Theorem 1.3. Fix $k \geq 1$ and let $c_1 = c_1(k)$ be as in Theorem 2.13. Let $\rho_1 > 0$, $\rho_2 : \mathbb{N} \to (0, 1]$, and f be as in Theorem 2.16 for k, and let $D = D(k)$ be so that $f(x) = x^{1/D}$ (see Theorem 2.16). Let $\mu_1 > 0$, $\mu_2 : \mathbb{N} \to (0, 1]$ be as in Proposition 2.12 for k. Set $\epsilon_1^* = \min\{\mu_1, \rho_1, (1/4)^D\}$ and define $\epsilon_2^* : \mathbb{N} \to (0, 1]$ by setting $\epsilon_2^*(x) = \min\{\mu_2(x), \rho_2(x), (1/2x)^{48}\}$, for each $x \in \mathbb{N}$.

Suppose $0 < \epsilon_1 < \epsilon_1^*$ and $\epsilon_2 : \mathbb{N} \to (0, 1]$ satisfies $\epsilon_2 < \epsilon_2^*$. We now choose a series of new constants. Set $\tau_1 = \epsilon_1^{4D}$ and note $\tau_1 < f(\epsilon_1)$. Set $\delta = \tau_1^{400}/1000$, $\epsilon_1' = (\delta/8c_1)^{2k+1000}$, $m = \lceil 2c_1(\delta/8)^{-2k-2}\rceil$, and $\epsilon_1'' = (\epsilon_1')^2/1000$. Next, define $\epsilon_2', \epsilon_2'' : \mathbb{N} \to (0, 1]$ by setting, for each $x \in \mathbb{N}$, $\epsilon_2'(x) = \epsilon_1''\epsilon_2(x)\epsilon_2(2^4\delta^{-8k-10})$ and $\epsilon_2''(x) = \epsilon_2(\delta^{-4}m^4)\epsilon_2'(x)^5/4$. Note there are polynomials $p_1(x)$, $p_2(x, y)$ depending only on k such that $\epsilon_1'' = p_1(\epsilon_1)$ and $\epsilon_2''(x) = p_2(\epsilon_1, x)$. To aid the reader in keeping track of the constants, we point out that the following inequalities hold:

$$\epsilon_1'' < \epsilon_1' < \delta < \tau_1 < \epsilon_1 < \epsilon_1^* \quad \text{and} \quad \epsilon_2'' < \epsilon_2' < \epsilon_2 < \epsilon_2^*.$$

Choose t_0 sufficiently large so that

$$\begin{aligned} \frac{t^3}{6} &\geq (1-\epsilon_1'')\binom{t}{3}, \\ \frac{(1-3\epsilon_1'')t^3}{12} &\geq (1-\epsilon_1')\binom{t}{3}, \text{ and} \\ \binom{t}{3}(1-6(\epsilon_1')^{1/4}-(\epsilon_1')^{3/8}) &\geq (1-\epsilon_1')^{1/8}\binom{t}{3}. \end{aligned}$$

Note there is some polynomial $p(x)$ depending only on k so that we can take $t_0 = p(\epsilon_1^{-1})$. Finally, choose T_1, L_1, and N_1 as in Theorem 2.7 for ϵ_1'', ϵ_2'', t_0 and $\ell_0 = 1$.

Set $L = \lceil \delta^{-4}m^4\rceil$, $T = T_1$, and choose N sufficiently large compared to all the previously chosen constants. Notice that $L = O_k(\epsilon_1^{-O_k(1)})$ and

$$T = T_0(p_1(\epsilon_1), \epsilon_2 \circ q_2, p(\epsilon_1^{-1}), 1),$$

where $T_0(x, y, z, w)$ is as in Theorem 2.7 and $q_2(y) = p_2(\epsilon_1, y)$.

Suppose $H = (V, E)$ is a 3-graph with $|V| \geq N$ satisfying $\mathrm{VC}_2(H) < k$. Theorem 2.16 implies there exist $1 \leq \ell \leq L_1$, $t_0 \leq t \leq T_1$, and $\mathcal{P}_1$ a $(t, \ell, \epsilon_1'', \epsilon_2''(\ell))$-decomposition of V which is $\mathrm{dev}_{2,3}(\epsilon_1'', \epsilon_2''(\ell_1))$-regular and $f(\epsilon_1'')$-homogeneous with respect to H. Say

$$\mathcal{P}_1 = \{V_1, \ldots, V_t\} \quad \text{and} \quad \mathcal{P}_2 = \left\{P_{ij}^{\alpha} : ij \in \binom{[t]}{2}, \alpha \in [\ell]\right\}.$$

Note that $f(\epsilon_1'') = (\epsilon_1'')^{1/D} < \frac{1}{4}$. Recall that as mentioned after Theorem 2.16, we may assume that all $\mathrm{dev}_{2,3}(\epsilon_1'', \epsilon_2''(\ell))$-regular triads of $\mathcal{P}$ are $f(\epsilon_1'')$-homogeneous with respect to H.

Given $ij \in \binom{[t]}{2}$ and $\alpha \in [\ell]$, let $G_{ij}^\alpha = (V_i \cup V_j, P_{ij}^\alpha)$. Given $ijs \in \binom{[t]}{3}$ and $1 \le \alpha, \beta, \gamma \le \ell$, set

$$G_{ijs}^{\alpha,\beta,\gamma} = (V_i \cup V_j \cup V_s, P_{ij}^\alpha \cup P_{js}^\beta \cup P_{is}^\gamma) \quad \text{and}$$
$$H_{ijs}^{\alpha,\beta,\gamma} = (V_i \cup V_j \cup V_s, E \cap K_3^{(2)}(G_{ijs}^{\alpha,\beta,\gamma})).$$

We will use throughout that since $\epsilon_2''(x) \le \epsilon_2'(x)^5/4$, Proposition 2.17 implies that for all $ijs \in \binom{[t]}{3}$ and $\alpha, \beta, \gamma \in [\ell]$,

$$|K_3^{(2)}(G_{ijs}^{\alpha,\beta,\gamma})| = (1 \pm \epsilon_2'(\ell))\left(\frac{n}{\ell t}\right)^3. \tag{1}$$

We use $\mathcal{P}$ to construct a different decomposition of V, which we call $\mathcal{Q}$, so that $\mathcal{Q}_1 = \mathcal{P}_1$ but $\mathcal{Q}_2 \ne \mathcal{P}_2$. Set

$$\begin{aligned}
\mathbf{F}_{\mathrm{err}} &= \{G_{ijs}^{\alpha,\beta,\gamma} \in \mathrm{Triads}(\mathcal{P}) : (H_{ijs}^{\alpha,\beta,\gamma}, G_{ijs}^{\alpha,\beta,\gamma}) \text{ fails } \mathrm{disc}_3(\epsilon_1'', \epsilon_2''(\ell))\},\\
\mathbf{F}_1 &= \{G_{ijs}^{\alpha,\beta,\gamma} \in \mathrm{Triads}(\mathcal{P}) \setminus \mathbf{F}_{\mathrm{err}} : d_{ijs}^{\alpha,\beta,\gamma} \ge 1 - f(\epsilon_1'')\}, \text{ and}\\
\mathbf{F}_0 &= \{G_{ijs}^{\alpha,\beta,\gamma} \in \mathrm{Triads}(\mathcal{P}) \setminus \mathbf{F}_{\mathrm{err}} : d_{ijs}^{\alpha,\beta,\gamma} \le f(\epsilon_1'')\}.
\end{aligned}$$

By assumption, $\mathrm{Triads}(\mathcal{P}) = \mathbf{F}_{\mathrm{err}} \sqcup \mathbf{F}_1 \sqcup \mathbf{F}_0$, and at most $\epsilon_1'' n^3$ triples $xyz \in \binom{V}{3}$ are in $K_3^{(2)}(G)$ for some $G \in \mathbf{F}_{\mathrm{err}}$. By (1), this implies

$$|\mathrm{Triads}(\mathcal{P}) \setminus \mathbf{F}_{\mathrm{err}}| \ge \left(\binom{n}{3} - \epsilon_1'' n^3\right) \Big/ \left(\frac{n^3}{t^3\ell^3}(1 - \epsilon_2'(\ell))\right) \ge \binom{t}{3}\ell^3(1 - \epsilon_1'),$$

where the last inequality uses that $t \ge t_0$ and n is large. Thus, $|\mathbf{F}_{\mathrm{err}}| \le \epsilon_1' t^3 \ell^3$. Let

$$\Psi = \left\{V_i V_j : \left|\{G_{ijs}^{\alpha\beta\gamma} \in \mathbf{F}_{\mathrm{err}} \text{ some } s \in [t] \text{ and } \alpha, \beta, \gamma \in [\ell]\}\right| \ge (\epsilon_1')^{3/4} \ell^3 t\right\}.$$

Since $|\mathbf{F}_{\mathrm{err}}| \le \epsilon_1' t^3 \ell^3$, we have that $|\Psi| \le (\epsilon_1')^{1/4} t^2$. Given $ij \in \binom{[t]}{2}$, let ℓ_{ij} be the number of $\alpha \in [\ell]$ such that G_{ij}^α has $\mathrm{dev}_2(\epsilon_2''(\ell), 1/\ell)$. After relabeling, we may assume $G_{ij}^1, \ldots, G_{ij}^{\ell_{ij}}$ each have $\mathrm{dev}_2(\epsilon_2''(\ell), 1/\ell)$. We claim that for $V_i V_j \notin \Psi$,

$$\ell_{ij} \ge (1 - 2(\epsilon_1')^{3/4})\ell.$$

Indeed, given $V_i V_j \notin \Psi$, if it were the case that $\ell_{ij} < (1 - 2(\epsilon_1')^{3/4})\ell$, then we would have that

$$\begin{aligned}
\left|\{G_{ijs}^{\alpha\beta\gamma} \in \mathbf{F}_{\mathrm{err}} \text{ some } s \in [t] \text{ and } \alpha, \beta, \gamma \in [\ell]\}\right| &\ge (t-2)\ell^2(\ell - \ell_{ij})\\
&> 2(\epsilon_1')^{3/4}(t-2)\ell^3 \ge (\epsilon_1')^{3/4}\ell^3 t,
\end{aligned}$$

contradicting $V_i V_j \notin \Psi$. Thus we have that for all $V_i V_j \notin \Psi$, $\ell_{ij} \ge (1 - 2(\epsilon_1')^{3/4})\ell$.

For each $V_iV_j \notin \Psi$, let H_{ij} be the edge-colored graph $(U_{ij} \cup W_{ij}, E^0_{ij}, E^1_{ij}, E^2_{ij})$, where

$$\begin{aligned}
W_{ij} &= \{P^\alpha_{ij} : \alpha \leq \ell_{ij}\},\\
U_{ij} &= \{P^\beta_{is}P^\gamma_{js} : s \in [t] \setminus \{i,j\}, \beta \leq \ell_{is}, \gamma \leq \ell_{js}\},\\
E^1_{ij} &= \{P^\alpha_{ij}(P^\beta_{is}P^\gamma_{js}) \in K_2[W_{ij}, U_{ij}] : G^{\alpha,\beta,\gamma}_{ijk} \in \mathbf{F}_1\},\\
E^0_{ij} &= \{P^\alpha_{ij}(P^\beta_{is}P^\gamma_{js}) \in K_2[W_{ij}, U_{ij}] : G^{\alpha,\beta,\gamma}_{ijk} \in \mathbf{F}_0\},\\
E^2_{ij} &= \{P^\alpha_{ij}(P^\beta_{is}P^\gamma_{js}) \in K_2[W_{ij}, U_{ij}] : G^{\alpha,\beta,\gamma}_{ijk} \in \mathbf{F}_{\mathrm{err}}\}.
\end{aligned}$$

By Proposition 2.12, and since $f(\epsilon''_1) < \frac{1}{2}$, H_{ij} contains no E^1_{ij}/E^0_{ij} copy of $U(k)$, and since $V_iV_j \notin \Psi$, $|E^2_{ij}| \leq (\epsilon'_1)^{3/4}\ell^3 t$. We will later need the following size estimates for W_{ij} and U_{ij}. By the above, $|W_{ij}| = \ell_{ij} \geq (1-2(\epsilon'_1)^{3/4})\ell$. We claim that $|U_{ij}| \geq (1-2(\epsilon'_1)^{3/4})\ell^2 t$. Indeed, observe that $|U_{ij}| = \sum_{s\in[t]\setminus\{i,j\}} \ell_{is}\ell_{js}$ and

$$\begin{aligned}
&\left|\{G^{\alpha\beta\gamma}_{ijs} \in \mathbf{F}_{\mathrm{err}} \text{ some } s \in [t] \text{ and } \alpha,\beta,\gamma \in [\ell]\}\right|\\
&\qquad\geq \sum_{s\in[t]\setminus\{i,j\}} \ell^2(\ell-\ell_{is}) + \ell_{is}\ell(\ell-\ell_{js})\\
&\qquad= \sum_{s\in[t]\setminus\{i,j\}} \ell^3 - \ell\ell_{is}\ell_{js} = (t-2)\ell^3 - \ell|U_{ij}|.
\end{aligned}$$

Since $V_iV_j \notin \Psi$, this shows that

$$(\epsilon'_1)^{3/4}\ell^3 t \geq (t-2)\ell^3 - \ell|U_{ij}|.$$

Rearranging, this yields that

$$|U_{ij}| \geq (t-2)\ell^2 - (\epsilon'_1)^{3/4}\ell^2 t \geq t\ell^2(1-2(\epsilon'_1)^{3/4}),$$

where the last inequality is because $t \geq t_0$.

Given $v, v' \in W_{ij}$, write $v \sim v' \in W_{ij}$ if for each $w \in \{0,1,2\}$,

$$|E^w_{ij}(v)\Delta E^w_{ij}(v')| \leq \delta|U_{ij}|.$$

By Lemma 2.14, there are $W^0_{ij} \subseteq W_{ij}$ of size at most $(\epsilon'_1)^{3/8}|W_{ij}|$, an integer $m_{ij} \leq m$, and $x^1_{ij}, \dots, x^{m_{ij}}_{ij} \in W_{ij}$ so that for all $v \in W_{ij} \setminus W^0_{ij}$, there is $1 \leq \alpha \leq m_{ij}$ so that $v \sim x^\alpha_{ij}$, and further, $|N_{E^2_{ij}}(v)| \leq (\epsilon'_1)^{3/8}|U_{ij}|$. For each $1 \leq u \leq m_{ij}$, let

$$W^u_{ij} = \{v \in W_{ij} \setminus W^0_{ij} : v \sim x^u_{ij} \text{ and for all } 1 \leq u' < u,\ v \nsim x^{u'}_{ij}\}.$$

Note $W^1_{ij} \cup \dots \cup W^{m_{ij}}_{ij}$ is a partition of $W_{ij} \setminus W^0_{ij}$.

We now define a series of sets to help us zero in on certain well behaved sets of triples. First, define

$$\Omega_0 = \left\{ ijs \in \binom{[t]}{3} : V_iV_j, V_jV_s, V_iV_s \notin \Psi \right\} \quad \text{and}$$
$$\Omega = \{W_{ij}^u W_{is}^v W_{js}^w : ijs \in \Omega_0, 1 \le u \le m_{ij}, 1 \le v \le m_{is}, 1 \le w \le m_{js}\}.$$

Since $|\Psi| \le (\epsilon_1')^{1/4} t^2$, $|\Omega_0| \ge \binom{t}{3} - |\Psi| t \ge (1 - 6(\epsilon_1')^{1/4})\binom{t}{3}$. Let

$$Y_0 = \bigcup_{W_{ij}^u W_{is}^v W_{js}^w \in \Omega} K_3[W_{ij}^u, W_{is}^v, W_{js}^w].$$

We have that for all $ijs \in \Omega_0$, $|W_{ij}^0|, |W_{is}^0|, |W_{js}^0| \le (\epsilon_1')^{3/8}\ell$, and therefore $|Y_0|$ is at least the following:

$$\begin{aligned} |Y_0| &\ge \binom{t}{3}\ell^3 - \ell^3 \left| \binom{[t]}{3} \setminus \Omega_0 \right| - |\Omega_0|(\epsilon_1')^{3/8}\ell^3 \\ &\ge \binom{t}{3}\ell^3 - 6(\epsilon_1')^{1/4}\binom{t}{3}\ell^3 - \binom{t}{3}(\epsilon_1')^{3/8}\ell^3 \\ &\ge \binom{t}{3}\ell^3(1 - (\epsilon_1')^{1/8}), \end{aligned}$$

where the last inequality is since $t \ge t_0$.

Given $ij \notin \Psi$, let us call W_{ij}^u *nontrivial* if it has size at least $\delta^{1/2}\ell/m_{ij}$. Define

$$\Omega_1 = \left\{ W_{ij}^u W_{js}^v W_{is}^w \in \Omega : \text{each of } W_{ij}^u, W_{js}^v, W_{is}^w \text{ are nontrivial} \right\},$$

and set $Y_1 = \bigcup_{W_{ij}^u W_{js}^v W_{is}^w \in \Omega_1} K_3[W_{ij}^u W_{js}^v W_{is}^w]$. Then we have that

$$\begin{aligned} |Y_1| &\ge |Y_0| - t\ell^2 \sum_{ij \in \binom{[t]}{2}} \sum_{\{u \in [m_{ij}] : W_{ij}^u \text{ trivial}\}} \delta^{1/2}(\ell/m_{ij}) \\ &\ge |Y_0| - t\ell^2(t^2\delta^{1/2}\ell) = |Y_0| - \delta^{1/2}t^3\ell^3. \end{aligned}$$

Define

$$\begin{aligned} R_1 &= \left\{ P_{ij}^\alpha P_{is}^\beta P_{js}^\gamma : G_{ijk}^{\alpha,\beta,\gamma} \in \mathbf{F}_1 \right\}, \\ R_0 &= \left\{ P_{ij}^\alpha P_{is}^\beta P_{js}^\gamma : G_{ijk}^{\alpha,\beta,\gamma} \in \mathbf{F}_0 \right\}, \\ R_2 &= \left\{ P_{ij}^\alpha P_{is}^\beta P_{js}^\gamma : G_{ijk}^{\alpha,\beta,\gamma} \in \mathbf{F}_{\text{err}} \right\}. \end{aligned}$$

Note that $(\mathcal{P}_2 \cup \mathcal{P}_2 \cup \mathcal{P}_2, R_0, R_1, R_2)$ is a 3-partite edge-colored 3-graph, and $|R_2| \le \epsilon_1' t^3 \ell^3$. Now set

$$\Omega_2 = \left\{ W_{ij}^u W_{js}^v W_{is}^w \in \Omega_1 : |R_2 \cap K_3[W_{ij}^u, W_{is}^v, W_{js}^w]| \le \sqrt{\epsilon_1'}\, |W_{ij}^u||W_{is}^v||W_{js}^w| \right\}$$

and $Y_2 = \bigcup_{W_{ij}^u W_{js}^v W_{is}^w \in \Omega_2} K_3[W_{ij}^u W_{js}^v W_{is}^w]$. Note that

$$\begin{aligned}|R_2| &\geq \sum_{W_{ij}^u W_{js}^v W_{is}^w \in \Omega_1 \setminus \Omega_2} \sqrt{\epsilon_1'}\, |W_{ij}^u||W_{is}^v||W_{js}^w| \\ &\geq \sqrt{\epsilon_1'} \sum_{W_{ij}^u W_{js}^v W_{is}^w \in \Omega_1 \setminus \Omega_2} |W_{ij}^u||W_{is}^v||W_{js}^w|.\end{aligned}$$

Therefore,

$$\sum_{W_{ij}^u W_{js}^v W_{is}^w \in \Omega_1 \setminus \Omega_2} |W_{ij}^u||W_{is}^v||W_{js}^w| \leq \sqrt{\epsilon_1'}^{-1} |R_2| < \sqrt{\epsilon_1'}^{-1} \epsilon_1' t^3 \ell^3 \leq \sqrt{\epsilon_1'}\, t^3 \ell^3.$$

This implies that $|Y_2| \geq |Y_1| - \sqrt{\epsilon_1'} t^3 \ell^3$.

Given $ijs \in \Omega_0$, let us call a triple $P_{ij}^\alpha P_{is}^\beta P_{js}^\gamma$ *troublesome* if one of the following hold:

- For some $u \in [m_{ij}]$, $P_{ij}^\alpha \in W_{ij}^u$, and there are $\sigma_1 \neq \sigma_2 \in \{0,1,2\}$ such that $P_{is}^\beta P_{js}^\gamma P_{ij}^\alpha \in R^{\sigma_1}$ and $P_{is}^\beta P_{js}^\gamma x_{ij}^u \in R^{\sigma_2}$.
- For some $w \in [m_{js}]$, $P_{js}^\gamma \in W_{js}^w$, and there are $\sigma_1 \neq \sigma_2 \in \{0,1,2\}$ such that $P_{is}^\beta P_{ij}^\alpha P_{js}^\gamma \in R^{\sigma_1}$ and $P_{is}^\beta P_{ij}^\alpha x_{js}^w \in R^{\sigma_2}$.
- For some $v \in [m_{is}]$, $P_{is}^\beta \in W_{is}^v$, and there are $\sigma_1 \neq \sigma_2 \in \{0,1,2\}$ such that $P_{ij}^\alpha P_{js}^\gamma P_{is}^\beta \in R^{\sigma_1}$ and $P_{ij}^\alpha P_{js}^\gamma x_{is}^v \in R^{\sigma_2}$.

Let Tr be the set of troublesome triples. Define

$$\Omega_3 = \big\{W_{ij}^u W_{js}^v W_{is}^w \in \Omega_2 : |K_3[W_{ij}^u W_{js}^v W_{is}^w] \cap \mathrm{Tr}| \leq \delta^{1/4} |W_{ij}^u||W_{js}^v||W_{is}^w|\big\},$$

and set $Y_3 = \bigcup_{W_{ij}^u W_{js}^v W_{is}^w \in \Omega_3} K_3[W_{ij}^u W_{js}^v W_{is}^w]$. We claim $|Y_3| \geq \binom{t}{3}\ell^3(1 - 2\delta^{1/2})$.

Given $V_i V_j \notin \Psi$, $1 \leq u \leq m_{ij}$, and $P_{ij}^\alpha \in W_{ij}^u$, we know that $P_{ij}^\alpha \sim x_{ij}^\alpha$, and therefore

$$\begin{aligned}&\big|\big\{P_{is}^\beta P_{js}^\gamma : s \in [t] \setminus \{i,j\}, \beta, \gamma \leq \ell, P_{is}^\beta P_{js}^\gamma P_{ij}^\alpha \in \mathrm{Tr}\big\}\big| \\ &\qquad\qquad \leq (\ell^2(t-2) - |U_{ij}|) + \sum_{x=0}^{2} |N_{E_{ij}^x}(P_{ij}^\alpha) \Delta N_{E_{ij}^x}(x_{ij}^u)| \\ &\qquad\qquad \leq 2(\epsilon_1')^{3/4} \ell^2 t + 3\delta t \ell^2 \\ &\qquad\qquad \leq 4\delta t \ell^2.\end{aligned}$$

Thus, $|\mathrm{Tr}| \leq 4\delta t\ell^2\big(\sum_{V_iV_j\notin\Psi, u\in[m_{ij}]}|W_{ij}^u|\big) \leq 4\delta t\ell^2(t^2\ell) = 4\delta t^3\ell^3$. Therefore

$$\begin{aligned}4\delta t^3\ell^3 \geq |\mathrm{Tr}| &\geq \sum_{W_{ij}^u W_{js}^v W_{is}^w\in\Omega_2\setminus\Omega_3} \delta^{1/4}|W_{ij}^u||W_{js}^v||W_{is}^w| \\ &= \delta^{1/4}\Bigg|\bigcup_{W_{ij}^u W_{js}^v W_{is}^w\in\Omega_2\setminus\Omega_3} K_3[W_{ij}^u W_{js}^v W_{is}^w]\Bigg|.\end{aligned}$$

Rearranging, this yields that

$$\Bigg|\bigcup_{W_{ij}^u W_{js}^v W_{is}^w\in\Omega_2\setminus\Omega_3} K_3[W_{ij}^u W_{js}^v W_{is}^w]\Bigg| \leq \delta^{-1/4}4\delta t^3\ell^3 = 4\delta^{3/4}t^3\ell^3.$$

Thus

$$\begin{aligned}|Y_3| &\geq |Y_2| - \delta^{3/4}t^3\ell^3 \\ &\geq |Y_1| - \sqrt{\epsilon_1'}t^3\ell^3 - 4\delta^{3/4}t^3\ell^3 \\ &\geq |Y_0| - \delta^{1/2}t^3\ell^3 - \sqrt{\epsilon_1'}t^3\ell^3 - 4\delta^{3/4}t^3\ell^3 \\ &\geq \binom{t}{3}\ell^3(1-7(\epsilon_1')^{1/8}) - \delta^{1/2}t^3\ell^3 - \sqrt{\epsilon_1'}t^3\ell^3 - 4\delta^{3/4}t^3\ell^3 \\ &\geq \binom{t}{3}\ell^3(1-2\delta^{1/2}).\end{aligned}$$

Therefore, using (1), we have

$$\begin{aligned}\Bigg|\bigcup_{P_{ij}^\alpha P_{is}^\beta P_{js}^\gamma\in Y_3} K_3^{(2)}(G_{ijs}^{\alpha,\beta,\gamma})\Bigg| &\geq \binom{t}{3}\ell^3(1-2\delta^{1/2})\Big(\frac{n^3}{t^3\ell^3}(1-\epsilon_2'(\ell))\Big) \\ &\geq \binom{n}{3}(1-3\delta^{1/2}),\end{aligned}$$

where the last inequality is because n is large.

Our next goal is to prove Claim 3.2, which says that for each $W_{ij}^u W_{is}^v W_{js}^w \in \Omega_3$, $K_3[W_{ij}^u, W_{is}^v, W_{js}^w]$ is either mostly contained in R_1 or mostly contained in R_0. For the proof of this claim, we will require the following notation. Given $ijs \in \binom{[t]}{3}$, $\alpha, \alpha' \leq \ell$, $1 \leq v \leq m_{is}$, and $1 \leq w \leq m_{js}$, we write $P_{ij}^\alpha \sim_{js,vw} P_{ij}^{\alpha'}$ if $P_{ij}^\alpha, P_{ij}^{\alpha'} \in W_{ij}^u$ for some $1 \leq u \leq m_{ij}$, and

$$\begin{aligned}\big|\{(P_{is}^\beta, P_{js}^\gamma) \in W_{is}^v \times W_{js}^w :& \text{ for some } \sigma_1 \neq \sigma_2 \in \{0,1,2\}, P_{is}^\beta P_{js}^\gamma P_{ij}^\alpha \in R^{\sigma_1} \\ &\text{and } P_{is}^\beta P_{js}^\gamma P_{ij}^{\alpha'} \in R^{\sigma_2}\}\big| \leq \delta^{1/8}|W_{is}^v||W_{js}^w|.\end{aligned}$$

Claim 3.2. *For any $W_{ij}^u W_{is}^v W_{js}^w \in \Omega_3$, there is $\sigma \in \{0,1\}$ such that*

$$\frac{|R_\sigma \cap K_3[W_{ij}^u, W_{is}^v, W_{js}^w]|}{|K_3[W_{ij}^u, W_{is}^v, W_{js}^w]|} \geq 1 - \delta^{1/100}.$$

Proof. Suppose towards a contradiction there is $W^u_{ij}W^v_{is}W^w_{js} \in \Omega_3$ such that for each $\sigma \in \{0, 1\}$,

$$\frac{|R_\sigma \cap K_3[W^u_{ij}, W^v_{is}, W^w_{js}]|}{|K_3[W^u_{ij}, W^v_{is}, W^w_{js}]|} < 1 - \delta^{1/100}.$$

To ease notation, let $A = W^u_{ij}$, $B = W^v_{is}$, and $C = W^w_{js}$.

We now define a series of subsets of A which will contain "well behaved" vertices. First, we set $A_1 = \{a \in A : a \sim_{js,vw} x^u_{ij}\}$. Since $W^u_{ij}W^v_{is}W^w_{js} \in \Omega_3$,

$$\delta^{1/4}|W^u_{ij}||W^v_{is}||W^w_{js}| \geq |\mathrm{Tr} \cap K_3[W^u_{ij}W^v_{is}W^w_{js}]| \geq |A \setminus A_1|\delta^{1/8}|W^v_{is}||W^w_{js}|.$$

Thus $|A \setminus A_1| \leq \delta^{-1/8}\delta^{1/4}|W^u_{ij}| = \delta^{1/8}|W^u_{ij}|$. Now set

$$A_2 = \{a \in A : |N_{R_2}(a)| \leq (\epsilon'_1)^{1/4}|B||C|\}.$$

Because $W^u_{ij}W^v_{is}W^w_{js} \in \Omega_2$, we have that

$$(\epsilon'_1)^{1/2}|A||B||C| \geq |R_2 \cap K_3[A, B, C]| \geq |A \setminus A_2|(\epsilon'_1)^{1/4}|B||C|.$$

Therefore, $|A \setminus A_2| \leq (\epsilon'_1)^{1/4}|A|$. Now set

$$A_3 = \{a \in A : |N_{R_1}(a)|/|B||C| \in (\delta^{1/64}, 1 - \delta^{1/64})\} \quad \text{and}$$
$$A'_3 = \{a \in A : |N_{R_1}(a)|/|B||C| \in (\delta^{1/128}, 1 - \delta^{1/128})\}.$$

We claim $x^u_{ij} \in A'_3$. Suppose towards a contradiction that $x^u_{ij} \notin A'_3$. Suppose first that $|N_{R_1}(x^u_{ij})| \geq (1 - \delta^{1/128})|B||C|$. Then for all $a \in A_1$, since $a \sim_{js,vw} x^u_{ij}$, we have $|N_{R_1}(a)| \geq (1 - \delta^{1/128} - \delta^{1/8})|B||C|$, and thus,

$$\begin{aligned}|R_1 \cap K_3[W^u_{ij}W^v_{is}W^w_{js}]| &\geq (1 - \delta^{1/128} - \delta^{1/8})|A_1||B||C| \\ &\geq (1 - \delta^{1/128} - \delta^{1/8})(1 - \delta^{1/8})|A||B||C| \\ &\geq (1 - \delta^{1/100})|A||B||C|,\end{aligned}$$

a contradiction. So we must have $|N_{R_1}(x^u_{ij})| \leq \delta^{1/128}|B||C|$. Then for all $a \in A_1 \cap A_2$, $a \sim_{js,vw} x^u_{ij}$ and $|N_{R_2}(a)| \leq (\epsilon'_1)^{1/4}|B||C|$ implies

$$|N_{R_0}(a)| \geq (1 - \delta^{1/128} - \delta^{1/8} - (\epsilon'_1)^{1/4})|B||C|.$$

Therefore

$$\begin{aligned}|R_0 \cap K_3[W^u_{ij}W^v_{is}W^w_{js}]| &\geq (1 - \delta^{1/128} - \delta^{1/8} - (\epsilon'_1)^{1/4})|A_1 \cap A_2||B||C| \\ &\geq (1 - \delta^{1/128} - \delta^{1/8} - (\epsilon'_1)^{1/4})(1 - \delta^{1/8} - (\epsilon'_1)^{1/4})|A||B||C| \\ &\geq (1 - \delta^{1/100})|A||B||C|,\end{aligned}$$

again a contradiction. Thus, we must have that $x^u_{ij} \in A'_3$. This implies that for all $a \in A_1 \cap A_2$,

$$\begin{aligned}|N_{R_1}(a)| &\geq |N_{R_1}(x^u_{ij})| - |N_{R_1}(x^u_{ij}) \Delta N_{R_1}(a)| \\ &\geq \delta^{1/128}|B||C|(1-\delta^{1/8}) \geq \delta^{1/64}|B||C|\end{aligned}$$

and

$$\begin{aligned}|N_{R_0}(a)| &\geq |N_{R_0}(x^u_{ij})| - |N_{R_0}(a)| - |N_{R_0}(x^u_{ij}) \Delta N_{R_0}(a)| \\ &\geq \delta^{1/128}|B||C|(1-\delta^{1/8} - (\epsilon'_1)^{1/4}) \\ &\geq \delta^{1/64}|B||C|.\end{aligned}$$

Thus $a \in A_3$. This shows that $A_1 \cap A_2 \subseteq A_3$, and therefore

$$|A_3| \geq |A|(1-\delta^{1/8} - (\epsilon'_1)^{1/4}).$$

Now define

$$A_B = \big\{a \in A : |\{b \in B : |N_{R_1}(ab) \Delta N_{R_1}(ax^v_{is})| \leq \delta^{1/16}|C|\}| \geq (1-\delta^{1/16})|B|\big\},$$
$$A_C = \big\{a \in A : |\{c \in C : |N_{R_1}(ac) \Delta N_{R_1}(ax^w_{js})| \leq \delta^{1/16}|B|\}| \geq (1-\delta^{1/16})|C|\big\}.$$

Observe that $4\delta^{1/4}|A||B||C| \geq |\mathrm{Tr} \cap K_3[A,B,C]| \geq \delta^{1/16}|A \setminus A_B||B||C|$, and therefore $|A \setminus A_B| \leq \delta^{-1/16} 4\delta^{1/4}|A| = 4\delta^{3/16}|A|$. A similar computation shows $|A \setminus A_C| \leq 4\delta^{3/16}|A|$. Consequently, setting $A_4 := A_3 \cap A_B \cap A_C$, we have that

$$|A_4| \geq |A_3| - |A \setminus A_C| - |A \setminus A_B| \geq |A|(1 - 8\delta^{3/16} - (\epsilon'_1)^{1/4} - \delta^{1/8}) > 0.$$

Fix some $a_* \in A_4$. We will use a_* to control the other edges in the triple. Let

$$S_1 = N_{R_1}(a^*), \quad S_0 = N_{R_0}(a^*), \quad \text{and} \quad S_2 = N_{R_2}(a^*).$$

Note $(B \cup C, S_0 \cup S_1 \cup S_2)$ is a 3-partite edge-colored 3-graph. Since $a_* \in A_3$, $|S_1|/|B||C| \in (\delta^{1/64}, 1-\delta^{1/64})$. Therefore, Corollary 2.19 implies that one of the following hold:

(a) There is $B_1 \subseteq B$ such that $|B_1| \geq \delta^{1/32}|B|/2$ and for all $b \in B_1$,

$$\frac{|N_{S_1}(b)|}{|C|} \in \left(\frac{\delta^{1/32}}{2}, 1 - \frac{\delta^{1/32}}{2}\right).$$

(b) There is $C_1 \subseteq C$ such that $|C_1| \geq \delta^{1/32}|C|/2$ and for all $c \in C_1$,

$$\frac{|N_{S_1}(c)|}{|B|} \in \left(\frac{\delta^{1/32}}{2}, 1 - \frac{\delta^{1/32}}{2}\right).$$

Without loss of generality, let us assume (a) holds (other case is symmetric). Define $B_2 = \{b \in B_1 : |N_{S_2}(b)| \leq (\epsilon'_1)^{1/16}|C|\}$. We claim $|B_2| \geq \delta^{1/32}|B|/4$. Indeed, we know that since $a_* \in A_2$,

$$(\epsilon'_1)^{1/4}|B||C| \geq |S_2| \geq (\epsilon'_1)^{1/16}|B_1 \setminus B_2||C|.$$

Thus, $|B_1 \setminus B_2| \leq (\epsilon_1')^{-1/16}(\epsilon_1')^{1/4}|B| = (\epsilon_1')^{1/12}|B|$, so

$$|B_2| \geq |B_1| - (\epsilon_1')^{1/12}|B| \geq \left(\frac{\delta^{1/32}}{2} - (\epsilon_1')^{1/12}\right)|B| \geq \frac{\delta^{1/32}|B|}{4}.$$

Note that for all $b \in B_2$, we have that $|N_{S_1}(b)| \geq \delta^{1/32}|C|/2 \geq \delta^{1/32}|C|/4$ and

$$|N_{S_0}(b)| \geq |C \setminus N_{S_1}(b)| - |N_{S_2}(b)| \geq \left(\frac{\delta^{1/32}}{2} - (\epsilon_1')^{1/16}\right)|C| \geq \frac{\delta^{1/32}|C|}{4}.$$

Now, let

$$B_3 = \{b \in B_2 : |N_{S_1}(b) \Delta N_{S_1}(x_{is}^v)| \leq \delta^{1/16}|C|\}.$$

Since $a_* \in A_B$,

$$|B_3| \geq |B_2| - \delta^{1/16}|B| \geq \left(\frac{\delta^{1/32}}{4} - \delta^{1/16}\right)|B| \geq \frac{\delta^{1/32}|B|}{8} > 0.$$

Fix some $b_* \in B_3$ and set $Q_0 = N_{S_0}(b_*)$ and $Q_1 = N_{S_1}(b_*)$. By above, since $b_* \in B_2$, $\min\{|Q_1|, |Q_0|\} \geq \delta^{1/32}|C|/4$.

We claim $|S_1 \cap K_2[B_3, Q_1]| \geq (1 - 10\delta^{1/32})|Q_1||B_3|$. Indeed, fix $b \in B_3$. Then we know $|N_{S_1}(b) \Delta N_{S_1}(x_{is}^v)| \leq \delta^{1/16}|C|$ and $|N_{S_1}(b_*) \Delta N_{S_1}(x_{is}^v)| \leq \delta^{1/16}|C|$, and therefore $|N_{S_1}(b) \Delta N_{S_1}(b_*)| \leq 2\delta^{1/16}|C|$. Consequently,

$$\begin{aligned}|N_{S_1}(b) \cap Q_1| &\geq |Q_1| - 2\delta^{1/16}|C| \geq |Q_1|\left(1 - 2\delta^{1/16}\frac{|C|}{|Q_1|}\right)\\ &\geq |Q_1|(1 - 2\delta^{1/16}(4\delta^{-1/32})) \geq |Q_1|(1 - 10\delta^{1/32}).\end{aligned}$$

This shows that $|S_1 \cap K_2[B_3, Q_1]| \geq (1 - 10\delta^{1/32})|Q_1||B_3|$.

Similarly, we claim $|S_0 \cap K_2[B_3, Q_0]| \geq (1 - 10\delta^{1/32})|B_3||Q_0|$. Indeed, for all $b \in B_3$, $|N_{S_2}(b)| \leq (\epsilon_1')^{1/16}|C|$ and, as above, $|N_{S_1}(b) \Delta N_{S_1}(b_*)| \leq 2\delta^{1/16}|C|$. Thus $|N_{S_0}(b) \Delta N_{S_0}(b_*)| \leq ((\epsilon_1')^{1/16} + 2\delta^{1/16})|C|$. Therefore,

$$\begin{aligned}|N_{S_0}(b) \cap Q_0| &\geq |Q_0| - ((\epsilon_1')^{1/4} + 2\delta^{1/16})|C|\\ &\geq |Q_0|\left(1 - ((\epsilon_1')^{1/4} + 2\delta^{1/16})\frac{|C|}{|Q_0|}\right)\\ &\geq |Q_0|(1 - ((\epsilon_1')^{1/4} + 2\delta^{1/16})4\delta^{-1/32})\\ &\geq |Q_1|(1 - 10\delta^{1/32}),\end{aligned}$$

where the last inequality uses the definition of ϵ_1'. This shows

$$|S_0 \cap K_2[B_3, Q_0]| \geq (1 - 10\delta^{1/32})|B_3||Q_0|.$$

Now let

$$\begin{aligned}Q_1' &= \{c \in Q_1 : |N_{S_1}(c) \cap B_3| \geq (1 - \sqrt{10}\,\delta^{1/64})|B_3|\} \quad \text{and}\\ Q_0' &= \{c \in Q_0 : |N_{S_0}(c) \cap B_3| \geq (1 - \sqrt{10}\,\delta^{1/64})|B_3|\}.\end{aligned}$$

Since both

$$|S_1 \cap K_2[B_3, Q_1]| \geq (1-10\delta^{1/32})|Q_1||B_3| \text{ and}$$
$$|S_0 \cap K_2[B_3, Q_0]| \geq (1-10\delta^{1/32})|B_3||Q_0|,$$

we have that $|Q_1'| \geq (1-\sqrt{10}\,\delta^{1/64})|Q_1|$ and $|Q_0'| \geq (1-\sqrt{10}\,\delta^{1/64})|Q_0|$. Finally, let

$$C^* = \{c \in C : |N_{S_1}(c)\Delta N_{S_1}(x_{js}^w)| \leq \delta^{1/16}|B|\}.$$

Since $a_* \in A_C$, $|C^*| \geq (1-\delta^{1/16})|C|$. Thus,

$$\begin{aligned}|Q_1' \cap C^*| &\geq (1-\sqrt{10}\,\delta^{1/64})|Q_1| - \delta^{1/16}|C| \\ &\geq (1-\sqrt{10}\delta^{1/64})\frac{\delta^{1/32}}{4} - \delta^{1/16}|C| \geq \frac{\delta^{1/32}|C|}{10}.\end{aligned}$$

Similarly,

$$\begin{aligned}|Q_0' \cap C^*| &\geq (1-\sqrt{10}\delta^{1/64})|Q_0| - \delta^{1/16}|C| \\ &\geq (1-\sqrt{10}\delta^{1/64})\frac{\delta^{1/32}}{4} - \delta^{1/16}|C| \geq \frac{\delta^{1/32}|C|}{10}.\end{aligned}$$

Consequently, there are $c_1 \in Q_1' \cap C^*$ and $c_0 \in Q_0' \cap C^*$. Since $c_0, c_1 \in C^*$, we can see that $|N_{S_1}(c_1)\Delta N_{S_1}(c_0)| \leq 2\delta^{1/16}|B|$. However, we also have that

$$\begin{aligned}|N_{S_1}(c_1) \cap N_{S_0}(c_0) \cap B_3| &\geq (1-2\sqrt{10}\,\delta^{1/64})|B_3| \\ &\geq (1-2\sqrt{10}\,\delta^{1/64})\delta^{1/32}\frac{|B|}{8} > 2\delta^{1/16}|B|.\end{aligned}$$

But this is a contradiction, since $N_{S_1}(c) \cap N_{S_0}(c_0) \cap B_3 \subseteq N_{S_1}(c_1)\Delta N_{S_1}(c_0)$. □

Let $\ell_1 = \lceil \delta^{-4}m^4 \rceil$. Suppose $V_iV_j \notin \Psi$ and $1 \leq u \leq \ell_{ij}$ is such that W_{ij}^u is nontrivial. Define $\mathbf{W}_{ij}^u = \bigcup_{P_{ij}^\alpha \in W_{ij}^u} P_{ij}^u$, let $\mathbf{G}_{ij}^u$ be the bipartite graph $(V_i \cup V_j, \mathbf{W}_{ij}^u)$, and define

$$\rho_{ij}(u) = \frac{|\mathbf{W}_{ij}^u|}{|V_i||V_j|}.$$

By Fact 2.23, $\mathbf{G}_{ij}^u$ has $\text{dev}_2(\ell(\epsilon_2''(\ell))^{1/4})$ and

$$|\mathbf{W}_{ij}^u| = (1 \pm \ell(\epsilon_2''(\ell))^{1/4})\frac{|W_{ij}^u||V_i||V_j|}{\ell}.$$

Using the size estimate above and the fact that W_{ij}^u is nontrivial, we have

$$\rho_{ij}(u) = (1 \pm \ell(\epsilon_2''(\ell))^{1/4})\frac{|W_{ij}^u|}{\ell} \geq (1 \pm \ell(\epsilon_2''(\ell))^{1/4})\frac{\delta^{1/2}}{\ell} \geq 2\ell(\epsilon_2''(\ell))^{1/4},$$

where the last inequality is by choice of $\epsilon_2''(\ell)$. Set $p_{ij}(u) = \rho_{ij}(u)^{-1}/\ell_1$, and let $s_{ij}(u) = [1/p_{ij}(u)]$. Observe that $\rho_{ij}p_{ij} = 1/\ell_1$. Note $(\epsilon_2''(\ell))^{1/4} \geq 10(1/s|V_i|)^{1/5}$

(since n is very large), and since W^u_{ij} is nontrivial and $\ell(\epsilon''_2(\ell))^{1/4} < \frac{1}{4}$,

$$\rho_{ij}(u) \geq (1 \pm \ell(\epsilon''_2(\ell))^{1/4})|W^u_{ij}|/\ell \geq \delta^{1/2}\frac{(\ell/m_{ij})}{\ell} = \delta^{1/2}/m_{ij} \geq \frac{\delta^{1/2}}{m}.$$

Further, $0 < p_{ij}(u) < \rho_{ij}(u)/2$ since

$$p_{ij}(u) \leq (1 \pm \ell(\epsilon''_2(\ell))^{1/4})^{-1} m\delta^{-1/2}\ell_1^{-1} \leq m\delta^{-1/2}\frac{\delta^4}{m^4} \leq \frac{\delta^{3/2}}{m^3} < \frac{\rho_{ij}(u)}{2},$$

where the last inequality uses that $\rho_{ij}(u) \geq \delta^{1/2}/m$. Thus by Lemma 2.22, there is a partition

$$\mathbf{W}^u_{ij} = \mathbf{W}^u_{ij}(0) \cup \cdots \cup \mathbf{W}^u_{ij}(s_{ij}(u)),$$

so that $|\mathbf{W}^u_{ij}(0)| \leq \rho_{ij} p_{ij}(1 + \epsilon'_1)|V_i||V_j|$ and for each $1 \leq x \leq s_{ij}(u)$, the bipartite graph $\mathbf{G}^u_{ij}(x) := (V_i \cup V_j, \mathbf{W}^u_{ij}(x))$ has $\mathrm{dev}_2(\ell(\epsilon''_2(\ell))^{1/4}, \rho_{ij} p_{ij})$, i.e., $\mathrm{dev}_2(\ell(\epsilon''_2(\ell))^{1/4}, 1/\ell_1)$. Since $(\epsilon''_2(\ell))^{1/4} m < \epsilon_2(\ell_1)$, and by definition of ϵ''_2, we have that for each $1 \leq x \leq s_{ij}(u)$, $\mathbf{G}^u_{ij}(x)$ has $\mathrm{dev}_2(\epsilon_2(\ell_1), 1/\ell)$. Let

$$s_{ij} = \sum_{1 \leq u \leq m_{ij}} s_{ij}(u).$$

Give a reenumeration

$$\{X^1_{ij}, \ldots, X^{s_{ij}}_{ij}\} = \{\mathbf{W}^u_{ij}(v) : 1 \leq v \leq s_{ij}(u), 1 \leq u \leq m_{ij}\}.$$

Then let $X^{s_{ij}+1}_{ij}, \ldots, X^{\ell_1}_{ij}$ be any partition of $K_2[V_i, V_j] \setminus \bigcup_{x=1}^{s_{ij}} X^x_{ij}$.

For $V_i V_j \in \Psi$ choose a partition $K_2[V_i, V_j] = X^1_{ij} \cup \cdots \cup X^{\ell_1}_{ij}$ such that for each $1 \leq x \leq \ell_1$, $X^{\ell_1}_{ij}$ has $\mathrm{dev}_2(\epsilon_2(\ell_1), 1/\ell_1)$ (such a partition exists by Lemma 2.22). Now define $\mathcal{Q}$ to be the decomposition of V with

$$\mathcal{Q}_1 = \{V_i : i \in [t]\} \quad \text{and} \quad \mathcal{Q}_2 = \left\{X^v_{ij} : v \leq \ell_1, ij \in \binom{[t]}{2}\right\}.$$

We claim this is a $(t, \ell_1, \epsilon_1, \epsilon_2(\ell_1))$-decomposition of V. Indeed, by construction, any $xy \in \binom{V}{2}$ which is not in an element of $\mathcal{Q}_2$ satisfying $\mathrm{disc}_2(\epsilon_2(\ell_1), 1/\ell_1)$ is in the set

$$\Gamma := \bigcup_{V_i V_j \notin \Psi} X^{s_{ij}+1}_{ij} \cup \cdots \cup X^{\ell_1}_{ij}.$$

Observe that

$$|\Gamma| \leq \sum_{V_i V_j \notin \Psi} \sum_{u=1}^{m_{ij}} |\mathbf{W}^u_{ij}(0)| + \left| K_2[V_i, V_j] \setminus \left(\bigcup_{P^\alpha_{ij} \in W_{ij}} P^\alpha_{ij} \right) \right|. \tag{2}$$

We have that

$$\begin{aligned}\sum_{V_iV_j\notin\Psi}\sum_{u=1}^{m_{ij}}|\mathbf{W}_{ij}^u(0)| &\le \sum_{V_iV_j\notin\Psi} m_{ij}(1+\epsilon_1')\rho_{ij}p_{ij}|V_i||V_j|\\ &\le \binom{t}{2}m(1+2\epsilon_1')\frac{(n/t)^2}{\ell_1}\\ &= \delta^4\binom{t}{2}(1+2\epsilon_1')\frac{(n/t)^2}{m^3}\le 2\delta^2\frac{\binom{n}{2}}{m^3},\end{aligned}$$

where the last inequality is because n is large. Then, by definition of δ and m, this shows that $\sum_{V_iV_j\notin\Psi}\sum_{u=1}^{m_{ij}}|\mathbf{W}_{ij}^u(0)|\le\epsilon_1\binom{n}{2}/2$. We also have that

$$\begin{aligned}&\sum_{V_iV_j\notin\Psi}\left|K_2[V_i,V_j]\setminus\left(\bigcup_{P_{ij}^\alpha\in W_{ij}}P_{ij}^\alpha\right)\right|\\ &\qquad\le \sum_{V_iV_j\notin\Psi}\left(|V_i||V_j|-|W_{ij}|(1+\epsilon_2'(\ell))\frac{|V_i||V_j|}{\ell}\right)\\ &\qquad= \sum_{V_iV_j\notin\Psi}|V_i||V_j|\left(1-\ell_{ij}(1+\epsilon_2'(\ell))\frac{1}{\ell}\right)\\ &\qquad\le \sum_{V_iV_j\notin\Psi}|V_i||V_j|\big(1-(1-2(\epsilon_1')^{3/4})(1+\epsilon_2'(\ell))\big)\\ &\qquad\le \sum_{V_iV_j\notin\Psi}|V_i||V_j|(\epsilon_1')^{1/8}\le(\epsilon_1')^{1/8}\binom{n}{2}.\end{aligned}$$

Combining these with (2) yields that $|\Gamma|\le\epsilon_1\binom{n}{2}/2+(\epsilon_1')^{1/8}\binom{n}{2}\le\epsilon_1\binom{n}{2}$, and therefore, $\mathcal{Q}$ is a $(t,\ell,\epsilon_1,\epsilon_2(\ell))$-decomposition of V.

We now show that $\mathcal{Q}$ is $\epsilon_1/6$-homogeneous with respect to H. We show first that for any $W_{ij}^uW_{is}^vW_{js}^w\in\Omega_3$, $\mathbf{G}_{ijs}^{uvw}:=(V_i\cup V_j\cup V_s,\mathbf{W}_{ij}^u\cup\mathbf{W}_{is}^v\cup\mathbf{W}_{js}^w)$ is $2\delta^{1/100}$-homogeneous with respect to H, and second that almost all $xyz\in K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})$ are in an $\epsilon_1/6$-homogenous triad of $\mathcal{Q}$.

Fix $W_{ij}^uW_{is}^vW_{js}^w\in\Omega_3$. We know by Claim 3.2, that there is $\sigma\in\{0,1\}$ such that

$$|R_\sigma\cap K_3[W_{ij}^u,W_{is}^v,W_{js}^w]|\ge(1-\delta^{1/100})|K_3[W_{ij}^u,W_{is}^v,W_{js}^w]|.$$

This implies, by (1) and definition of R_σ, that the following holds, where $E^1=E$ and $E^0=\binom{V}{3}\setminus E^1$ (recall $E=E(H)$):

$$\begin{aligned}&|E^\sigma\cap K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|\\ &\qquad\ge(1-\delta^{1/100})(1-\epsilon_1'')|K_3[W_{ij}^u,W_{is}^v,W_{js}^w]|(1-\ell^3\epsilon_2'(\ell))|V_i||V_j||V_s|\frac{1}{\ell^3}\\ &\qquad=(1-\delta^{1/100})(1-\epsilon_1'')(1-\ell^3\epsilon_2'(\ell))\frac{|W_{ij}^u||W_{is}^v||W_{js}^w|}{\ell^3}\end{aligned}$$

On the other hand, note that by (1),

$$|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})| = |W_{ij}^u||W_{is}^v||W_{js}^w|(1 \pm \ell^3\epsilon_2'(\ell))\frac{|V_i||V_j||V_s|}{\ell^3}.$$

Combining this with the above, we see that

$$\begin{aligned}|E^\sigma \cap K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|&\\ &\geq (1-\delta^{1/100})(1-\epsilon_1'')(1-\ell^3\epsilon_2'(\ell))(1+\ell^3\epsilon_2'(\ell))^{-1}|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|\\ &\geq (1-2\delta^{1/100})|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|,\end{aligned}$$

where the last inequality is by definition of ϵ_2' and ϵ_1''. This shows $\mathbf{G}_{ijs}^{uvw}$ is $2\delta^{1/100}$-homogeneous. We now show that almost all $xyz \in K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})$ are in an $\epsilon_1/6$-homogeneous triad of $\mathcal{Q}$. Set

$$\Sigma_0(ijs, uvw) = \{0, \dots, s_{ij}(u)\} \times \{0, \dots, s_{is}(v)\} \times \{0, \dots, s_{js}(w)\}.$$

Given $(x, y, z) \in \Sigma_0$, set

$$\mathbf{G}_{ijs}^{uvw}(x, y, z) = (V_i \cup V_j \cup V_s; \mathbf{W}_{ij}^u(x) \cup \mathbf{W}_{is}^v(y) \cup \mathbf{W}_{js}^w(z)).$$

Note that $K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}) = \bigcup_{(x,y,z)\in\Sigma_0(ijs,uvw)} K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))$. Define

$$\begin{aligned}&\Sigma_1(ijs, uvw)\\ &\quad = \{(x, y, z) \in \{0, \dots, s_{ij}(u)\} \times \{0, \dots, s_{is}(v)\} \times \{0, \dots, s_{js}(w)\} : x, y \text{ or } z \text{ is } 0\},\end{aligned}$$

and set $\Sigma_2(ijs, uvw) = \Sigma_0(ijs, uvw) \setminus \Sigma_1(ijs, uvw)$. Note that by construction, for all $(x, y, z) \in \Sigma_2(ijs, uvw)$, $\mathbf{G}_{ijs}^{uvw}(x, y, z) \in \text{Triads}(\mathcal{Q})$. Observe that

$$\begin{aligned}\sum_{(x,y,z)\in\Sigma_1(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))|&\\ &\leq |\mathbf{W}_{ij}^u(0)||V_s| + |\mathbf{W}_{is}^v(0)||V_j| + |\mathbf{W}_{js}^w(0)||V_i|\\ &\leq (1+\epsilon_1')|V_i||V_j||V_s|(\rho_{ij}p_{ij} + \rho_{is}p_{is} + \rho_{js}p_{js})\\ &= 3(1+\epsilon_1')|V_i||V_j||V_s|\frac{1}{\ell_1}\\ &\leq 3(1+\epsilon_1')\delta^4|V_i||V_j||V_s|m^{-4}\\ &\leq 3(1+\epsilon_1')\delta^4\Big(|W_{ij}^u||W_{is}^v||W_{js}^w|\frac{1}{\ell^3}\Big)^{-1}m^{-4}|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|\\ &\leq 3(1+\epsilon_1')\delta^4\Big(\frac{\delta^{1/2}}{m}\Big)^{-3}m^{-4}|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|\\ &= 3(1+\epsilon_1')\delta^{1/2}m^{-1}|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|\\ &< \delta|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|,\end{aligned}$$

where the last inequality uses the definition of m. Let $\Sigma_3(ijs, uvw)$ be the set of $(x, y, z) \in \Sigma_2(ijs, uvw)$ such that

$$|E^\sigma \cap K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| < (1-\delta^{1/200})|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))|,$$

and set

$$\Sigma_4(ijs, uvw) = \Sigma_2(ijs, uvw) \setminus \Sigma_3(ijs, uvw).$$

By definition, and since $\delta^{1/200} < \epsilon_1/6$, every triad of the form $K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))$ for $(x, y, z) \in \Sigma_4(ijs, uvw)$ is in an $\epsilon_1/6$-homogeneous triad of $\mathcal{Q}$. We now show that $\bigcup_{(x,y,z)\in\Sigma_4(ijs,uvw)} K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})$ is most of $K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})$. Observe

$$\begin{aligned}
|E^\sigma \cap K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})| \le & \sum_{(x,y,z)\in\Sigma_1(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| \\
& + (1-\delta^{1/200}) \sum_{(x,y,z)\in\Sigma_3(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| \\
& + \sum_{(x,y,z)\in\Sigma_4(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| \\
\le & \,\delta |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})| \\
& + (1-\delta^{1/200}) \sum_{(x,y,z)\in\Sigma_3(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| \\
& + \sum_{(x,y,z)\in\Sigma_4(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))|.
\end{aligned}$$

Thus, since $|E^\sigma \cap K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})| \ge (1-2\delta^{1/100})|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})|$,

$$\begin{aligned}
&(1-2\delta^{1/100}-\delta)|K_3^{(2)}(\mathbf{G}_{ijs}^{uvw})| \\
&\quad \le (1-\delta^{1/200}) \sum_{(x,y,z)\in\Sigma_3(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| \\
&\qquad + \sum_{(x,y,z)\in\Sigma_4(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| \\
&\quad = \sum_{(x,y,z)\in\Sigma_2(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))| \\
&\qquad - \delta^{1/200} \sum_{(x,y,z)\in\Sigma_3(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))|.
\end{aligned}$$

Rearranging this inequality, we have the following upper bound for the sum $\sum_{(x,y,z)\in\Sigma_3(ijs,uvw)} |K_3^{(2)}(\mathbf{G}_{ijs}^{uvw}(x, y, z))|$:

$$\delta^{-1/200}\Bigg(\sum_{(x,y,z)\in\Sigma_2(ijs,uvw)}|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs}(x,y,z))|-(1-2\delta^{1/100}-\delta)|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})|\Bigg)$$
$$\leq \delta^{-1/200}|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})|3\delta^{1/100}$$
$$\leq 3\delta^{1/200}|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})|.$$

Consequently,

$$\sum_{(x,y,z)\in\Sigma_4(ijs,uvw)}|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs}(x,y,z))|\geq |K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})|(1-3\delta^{1/200}).$$

We have now established that $\bigcup_{(x,y,z)\in\Sigma_4(ijs,uvw)}K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})$ covers most of $K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})$, and for all $(x,y,z)\in\Sigma_4(ijs,uvw)$, $\mathbf{G}^{uvw}_{ijs}(x,y,z)$ is an $\epsilon_1/6$-homogeneous triad of $\mathcal{Q}$. For all $(x,y,z)\in\Sigma_4(ijs,uvw)$, $W^u_{ij}(x)$, $W^v_{is}(y)$, $W^w_{js}(z)$ all have $\mathrm{dev}_2(\epsilon_2(\ell_1),1/\ell_1)$, and thus, by Proposition 2.24, $\mathbf{G}^{uvw}_{ijs}(x,y,z)$ has $\mathrm{dev}_{2,3}(\epsilon_1,\epsilon_2(\ell_1))$ with respect to H.

Using this and our lower bound on the size of Y_3, we can now give the following lower bound on the number of triples $xyz\in\binom{V}{3}$ in a $\mathrm{dev}_{2,3}(\epsilon_2(\ell_1),\epsilon_1)$-regular triad of $\mathcal{P}$:

$$\sum_{W^u_{ij}W^v_{is}W^w_{js}\in\Omega_3}\;\sum_{(x,y,z)\in\Sigma_4(ijs,uvw)}|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs}(x,y,z))|$$
$$\geq \sum_{W^u_{ij}W^v_{is}W^w_{js}\in\Omega_3}(1-3\delta^{1/200})|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})|$$
$$=(1-3\delta^{1/200})\sum_{W^u_{ij}W^v_{is}W^w_{js}\in\Omega_3}|K_3^{(2)}(\mathbf{G}^{uvw}_{ijs})|$$
$$\geq(1-3\delta^{1/200})(1-3\delta^{1/2})\binom{n}{3}$$
$$\geq(1-\epsilon_1)\binom{n}{3},$$

where the last inequality is by definition of δ. This finishes the proof. □

Appendix: Proof of Proposition 2.24

We will use the following fact:

Lemma A.1. *For all $\delta,r,\mu\in(0,1]$ satisfying $2^{12}\delta<\mu^2r^{12}$, the following holds. Suppose $G=(V_1\cup V_2\cup V_3,E)$ is a 3-partite graph such that for each $ij\in\binom{[3]}{2}$, $||V_i|-|V_j||\leq\delta|V_i|$ and $G[V_i,V_j]$ has $\mathrm{dev}_2(\delta,r)$. Given $u_0v_0w_0\in K_3^{(2)}(G)$, define*

$$K_{2,2,2}[u_0,v_0,w_0]$$
$$=\big\{u_1v_1w_1\in K_3[V_1,V_2,V_3]:\textit{ for each }\epsilon\in\{0,1\}^3,(u_{\epsilon_1},v_{\epsilon_2},w_{\epsilon_3})\in K_3^{(2)}(G)\big\}.$$

Then if $J := \{uvw \in K_3^{(2)}(G) : |K_{2,2,2}[u_0, v_0, w_0]| \le (1+\mu)r^9|V_1||V_2||V_3|\}$, *we have that* $|J| \ge (1-\mu)r^3|V_1||V_2||V_3|$.

Proof. Let $K^G_{2,2,2}[V_1, V_2, V_3]$ be the set

$$\{(u_0, u_1, w_0, w_1, z_0, z_1) \in V_1^2 \times V_2^2 \times V_3^3 \\ : \text{for each } \epsilon \in \{0,1\}^3, u_{\epsilon_1} w_{\epsilon_2} z_{\epsilon_3} \in R \cap K_3^{(2)}(G)\}.$$

By Theorem 3.5 in [11],

$$|K_{2,2,2}[V_1, V_2, V_3]| \le r^{12}|V_1|^2|V_2|^2|V_3|^2 + 2^{12}\delta^{1/4}|V_1|^2|V_2|^2|V_3|^2.$$

Suppose towards a contradiction that $|J| > (1-\mu)r^3|V_1||V_2||V_3|$. Then

$$|K_{2,2,2}[V_1, V_2, V_3]| \ge |J|(1+\mu)r^9|V_1||V_2||V_3| > (1-\mu^2)r^{12}|V_1|^2|V_2|^2|V_3|^2.$$

Combining with the above, this implies $r^{12} + 2^{12}\delta^{1/4} > (1-\mu^2)r^{12}$, which implies $\mu^2 r^{12} < 2^{12}\delta^{1/4}$, a contradiction. ☐

Proof of Proposition 2.24. Fix $0 < \epsilon < \frac{1}{2}$, $0 < d_2 < \frac{1}{2}$, and $0 < \delta \le (d_2/2)^{48}$, and choose N sufficiently large.

Suppose $H = (V_1 \cup V_2 \cup V_3, R)$ is a 3-partite 3-graph on $n \ge N$ vertices and for each $i, j \in [3]$, $||V_i| - |V_j|| \le \delta|V_i|$. Suppose $G = (V_1 \cup V_2 \cup V_3, E)$ is a 3-partite graph, where for each $1 \le i < j \le 3$, $G[V_i, V_j]$ has $\mathrm{dev}_2(\delta, d_2)$, and assume $|R \cap K_3^{(2)}(G)| \le \epsilon|K_3^{(2)}(G)|$. Let d be such that $|R \cap K_3^{(2)}(G)| = d|K_3^{(2)}(G)|$. By assumption $d \le \epsilon$. Define $g(x, y, z) : \binom{V}{3} \to [0, 1]$ by

$$g(x, y, z) = \begin{cases} 1-d & \text{if } xyz \in R \cap K_3^{(2)}(G), \\ -d & \text{if } xyz \in K_3^{(2)}(G) \setminus R, \\ 0 & \text{otherwise.} \end{cases}$$

Given $u_0v_0w_0 \in K_3^{(2)}(G)$, define

$$K_{2,2,2}[u_0, v_0, w_0] = \{u_1v_1w_1 \in K_3[V_1, V_2, V_3] \\ : \text{for each } (i, j, k) \in \{0,1\}^3, (u_i, v_j, w_k) \in K_3^{(2)}(G)\}.$$

Let $\mu = d_2^{12}$. Note that $2^{12}\delta < (d_2/2)^{36} < d_2^{36} = \mu^2 d_2^{12}$. Set

$$J := \big\{uvw \in K_3^{(2)}(G) : |K_{2,2,2}[u_0, v_0, w_0]| \le (1+\mu)d_2^9|V_1||V_2||V_3|\big\}.$$

By Lemma A.1, we have that $|J| \ge (1-\mu)d_2^3|V_1||V_2||V_3|$. Now set

$$I_1 = \big\{(u_0, u_1, w_0, w_1, z_0, z_1) \in V_1^2 \times V_2^2 \times V_3^3 \\ : \text{for each } (i, j, k) \in \{0,1\}^3, u_i w_j z_k \in R \cap K_3^{(2)}(G)\big\}$$

and let

$$I_2 = \{(u_0, u_1, w_0, w_1, z_0, z_1) \in (V_1^2 \times V_2^2 \times V_3^3) \setminus I_1 \\ : \text{for each } (i, j, k) \in \{0,1\}^3, u_i w_j z_k \in K_3^{(2)}(G)\}.$$

Then

$$\begin{aligned}\sum_{u_0,u_1\in V_1} \sum_{w_0,w_1\in V_2} \sum_{z_0,z_1\in V_3} \prod_{(i,j,k)\in\{0,1\}^3} & g(u_i, w_j, z_k) \\ &\leq \left| \sum_{u_0,u_1\in V_1} \sum_{w_0,w_1\in V_2} \sum_{z_0,z_1\in V_3} \prod_{(i,j,k)\in\{0,1\}^3} g(u_i, w_j, z_k) \right| \\ &\leq \sum_{u_0,u_1\in V_1} \sum_{w_0,w_1\in V_2} \sum_{z_0,z_1\in V_3} \left| \prod_{(i,j,k)\in\{0,1\}^3} g(u_i, w_j, z_k) \right| \\ &= \sum_{(u_0,u_1,w_0,w_1,z_0,z_1)\in I_1} (1-d)^9 \\ &\qquad + \sum_{(u_0,u_1,w_0,w_1,z_0,z_1)\in I_2} \left| \prod_{(i,j,k)\in\{0,1\}^3} g(u_i, w_j, z_k) \right|.\end{aligned}$$

For each $(u_0, u_1, w_0, w_1, z_0, z_1) \in I_2$,

$$\left| \prod_{(i,j,k)\in\{0,1\}^3} g(u_i, w_j, z_k) \right| \leq d(1-d)^8,$$

since at least one of the $g(u_i, w_j, z_k)$ is equal to $-d$, and $|-d| < |1-d|$ (since $d \leq \epsilon < \frac{1}{2}$). Thus we have, by above, that

$$\sum_{u_0,u_1\in V_1} \sum_{w_0,w_1\in V_2} \sum_{z_0,z_1\in V_3} \prod_{(i,j,k)\in\{0,1\}^3} g(u_i, w_j, z_k) \leq (1-d)^9 |I_1| + d(1-d)^8 |I_2|.$$

Note

$$\begin{aligned}|I_1| &\leq \sum_{u_0 w_0 z_0 \in J} |K_{2,2,2}(u_0, w_0, z_0)| + \sum_{u_0 w_0 z_0 \in R\setminus J} |K_{2,2,2}(u_0, w_0, z_0)| \\ &\leq |J|(1+\mu) d_2^9 |V_1||V_2||V_3| + |R \setminus J||R| \\ &\leq |R|(1+\mu) d_2^9 |V_1||V_2||V_3| + \mu d_2^3 |V_1||V_2||V_3| d |K_3^{(2)}(G)| \\ &\leq d|K_3^{(2)}(G)|(1+\mu) d_2^9 |V_1||V_2||V_3| + \mu d_2^3 |V_1||V_2||V_3| d |K_3^{(2)}(G)| \\ &\leq |V_1||V_2||V_3||K_3^{(2)}(G)|(d(1+\mu) d_2^{12} + d d_2^{12}),\end{aligned}$$

where the last inequality is by definition of μ. By the counting lemma [11, Theorem 3.5], $|K_3^{(2)}(G)| \leq (1+2^3\delta^{1/4})|V_1||V_2||V_3|$. Therefore, we have that

$$|I_1| \leq |V_1|^2|V_2|^2|V_3|^2(1+2^3\delta^{1/4})(d(1+\mu)d_2^{12}+dd_2^{12}) \leq 3dd_2^{12}|V_1|^2|V_2|^2|V_3|^2.$$

On the other hand, $|I_2| \leq |K_{2,2,2}[V_2, V_2, V_3]|$, which, by [11, Theorem 3.5], has size at most $(d_2^{12}+2^{12}\delta^{1/4})|V_1|^2|V_2|^2|V_3|^2$. Combining the bounds above with the fact that $d \leq \epsilon$, we have that

$$\begin{aligned}\sum_{u_0,u_1\in V_1}\sum_{w_0,w_1\in V_2}\sum_{z_0,z_1\in V_3}\prod_{(i,j,k)\in\{0,1\}^3} & g(u_i,w_j,z_k) \\ &\leq (1-d)^9|I_1| + d(1-d)^8|I_2| \\ &\leq |V_1|^2|V_2|^2|V_3|^2(3\epsilon d_2^{12} + \epsilon(d_2^{12}+2^{12}\delta^{1/4})) \\ &\leq 6\epsilon d_2^{12}|V_1|^2|V_2|^2|V_3|^2,\end{aligned}$$

where the last inequality is due to $\delta < (d_2/2)^{48}$. This shows that (H, G) has $\mathrm{dev}_{2,3}(\delta, 6\epsilon)$, as required. □

References

[1] N. Alon, E. Fischer, and I. Newman, "Efficient testing of bipartite graphs for forbidden induced subgraphs", *SIAM J. Comput.* **37**:3 (2007), 959–976. MR Zbl

[2] A. Chernikov and N. Hempel, "On n-dependent groups and fields, II", *Forum Math. Sigma* **9** (2021), art. id. e38. MR Zbl

[3] A. Chernikov and S. Starchenko, "Definable regularity lemmas for NIP hypergraphs", *Q. J. Math.* **72**:4 (2021), 1401–1433. MR Zbl

[4] A. Chernikov and H. Towsner, "Hypergraph regularity and higher arity VC-dimension", preprint, 2020. arXiv 2010.00726

[5] A. Chernikov, D. Palacin, and K. Takeuchi, "On n-dependence", *Notre Dame J. Form. Log.* **60**:2 (2019), 195–214. MR Zbl

[6] D. Conlon and J. Fox, "Bounds for graph regularity and removal lemmas", *Geom. Funct. Anal.* **22**:5 (2012), 1191–1256. MR Zbl

[7] J. Fox and L. M. Lovász, "A tight lower bound for Szemerédi's regularity lemma", *Combinatorica* **37**:5 (2017), 911–951. MR Zbl

[8] J. Fox, J. Pach, and A. Suk, "Erdős–Hajnal conjecture for graphs with bounded VC-dimension", *Discrete Comput. Geom.* **61**:4 (2019), 809–829. MR Zbl

[9] P. Frankl and V. Rödl, "Extremal problems on set systems", *Random Structures Algorithms* **20**:2 (2002), 131–164. MR Zbl

[10] W. T. Gowers, "Lower bounds of tower type for Szemerédi's uniformity lemma", *Geom. Funct. Anal.* **7**:2 (1997), 322–337. MR Zbl

[11] W. T. Gowers, "Quasirandomness, counting and regularity for 3-uniform hypergraphs", *Combin. Probab. Comput.* **15**:1-2 (2006), 143–184. MR Zbl

[12] W. T. Gowers, "Hypergraph regularity and the multidimensional Szemerédi theorem", *Ann. of Math.* (2) **166**:3 (2007), 897–946. MR Zbl

[13] D. Haussler, "Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik–Chervonenkis dimension", *J. Combin. Theory Ser. A* **69**:2 (1995), 217–232. MR Zbl

[14] P. E. Haxell, B. Nagle, and V. Rödl, "An algorithmic version of the hypergraph regularity method", pp. 439–448 in *Proceedings of the* 46*th Annual IEEE Symposium on Foundations of Computer Science* (Pittsburgh, PA, 2005), IEEE Computer Society Press, Washington, DC, 2005.

[15] N. Hempel, "On n-dependent groups and fields", *MLQ Math. Log. Q.* **62**:3 (2016), 215–224. MR Zbl

[16] L. Lovász and B. Szegedy, "Regularity partitions and the topology of graphons", pp. 415–446 in *An irregular mind*, edited by I. Bárány and J. Solymosi, Bolyai Soc. Math. Stud. **21**, János Bolyai Math. Soc., Budapest, 2010. MR Zbl

[17] G. Moshkovitz and A. Shapira, "A short proof of Gowers' lower bound for the regularity lemma", *Combinatorica* **36**:2 (2016), 187–194. MR Zbl

[18] G. Moshkovitz and A. Shapira, "A tight bound for hyperaph regularity", *Geom. Funct. Anal.* **29**:5 (2019), 1531–1578. MR Zbl

[19] B. Nagle, A. Poerschke, V. Rödl, and M. Schacht, "Hypergraph regularity and quasi-randomness", pp. 227–235 in *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*, edited by C. Mathieu, SIAM, Philadelphia, PA, 2009. MR Zbl

[20] V. Rödl and J. Skokan, "Regularity lemma for k-uniform hypergraphs", *Random Structures Algorithms* **25**:1 (2004), 1–42. MR Zbl

[21] V. Rödl and J. Skokan, "Counting subgraphs in quasi-random 4-uniform hypergraphs", *Random Structures Algorithms* **26**:1-2 (2005), 160–203. MR Zbl

[22] S. Shelah, "Strongly dependent theories", *Israel J. Math.* **204**:1 (2014), 1–83. MR Zbl

[23] S. Shelah, "Definable groups for dependent and 2-dependent theories", *Sarajevo J. Math.* **13(25)**:1 (2017), 3–25. MR Zbl

[24] E. Szemerédi, "Regular partitions of graphs", pp. 399–401 in *Problèmes combinatoires et théorie des graphes* (Orsay, France, 1976), Colloq. Internat. CNRS **260**, CNRS, Paris, 1978. MR Zbl

[25] C. Terry, "VC_ℓ-dimension and the jump to the fastest speed of a hereditary $\mathcal{L}$-property", *Proc. Amer. Math. Soc.* **146**:7 (2018), 3111–3126. MR Zbl

[26] C. Terry and J. Wolf, "Irregular triads in 3-uniform hypergraphs", preprint, 2021. arXiv 2111.01737

Received 11 Apr 2022. Revised 1 Nov 2022.

CAROLINE TERRY:

terry.376@osu.edu
Department of Mathematics, The Ohio State University, Columbus, OH, United States

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.357

Galois groups of large simple fields

Anand Pillay and Erik Walsberg
Appendix by Philip Dittmann

For Ehud Hrushovski, on his 60th birthday.

Suppose that K is an infinite field which is large (in the sense of Pop) and whose first-order theory is simple. We show that K is *bounded*, namely has only finitely many separable extensions of any given finite degree. We also show that any genus 0 curve over K has a K-point and if K is additionally perfect then K has trivial Brauer group. These results give evidence towards the conjecture that large simple fields are bounded PAC. Combining our results with a theorem of Lubotzky and van den Dries we show that there is a bounded PAC field L with the same absolute Galois group as K. In the appendix we show that if K is large and NSOP_∞ and v is a nontrivial valuation on K then (K, v) has separably closed Henselization, so in particular the residue field of (K, v) is algebraically closed and the value group is divisible. The appendix also shows that formally real and formally p-adic fields are SOP_∞ (without assuming largeness).

1. Introduction

Throughout K is a field. Large fields were introduced by Pop [1996], one definition being that K is *large* if any algebraic curve defined over K with a smooth (nonsingular) K-point has infinitely many K-points. Finite fields, number fields, and function fields are not large. Local fields, Henselian fields, quotient fields of Henselian domains, real closed fields, separably closed fields, pseudofinite fields, infinite algebraic extensions of finite fields, and fields which satisfy a local-global principle (in particular pseudo-real closed and pseudo-p-adically closed fields) are all large. All infinite fields whose first-order theory is known to be "tame" or well-behaved in various senses are large. Let K^{sep} be a separable closure of K. We say that K is *bounded* if for any n there are only finitely many degree n extensions of K in K^{sep}, or equivalently if the absolute Galois group $\mathrm{Aut}(K^{\mathrm{sep}}/K)$ of K has only finitely many open subgroups of any given finite index. When K is also perfect, this is also called Serre's property (F). (Other authors use "bounded" to mean that

MSC2020: 03C45, 03C60, 12F10, 12J10.
Keywords: large field, bounded field, simple theory, stable theory.

K has only finitely many extensions of each degree.) Koenigsmann has conjectured that bounded fields are large [Junker and Koenigsmann 2010, p. 496].

Recall that K is *pseudoalgebraically closed* (PAC) if any geometrically integral K-variety V has a K-point (and hence the set $V(K)$ of K-points is Zariski dense in V). We mention in passing that a PAC field need not be perfect; if $p = \mathrm{Char}(K)$, and $a \in K$ is not a p-th power, then Spec $K[x]/(x^p - a)$ is not geometrically integral [Poonen 2017, Example 2.2.9]. PAC fields are large, by definition. PAC fields were introduced by Ax [1968], who showed that pseudofinite fields are bounded PAC. Infinite algebraic extensions of finite fields are also bounded PAC [Fried and Jarden 2005, Corollary 11.2.4]. In either case PAC follows from the Hasse–Weil estimates.

On the model-theoretic side, we have various "tame" classes of first-order theories T, the most "perfect" being stable theories, and some others being simple theories and NIP theories. It is a well-known theorem of Shelah that a theory is stable if and only if it is both simple and NIP. Good examples come from theories of fields. We say that a first-order structure, in particular a field, is stable (simple, NIP) if its theory is stable (simple, NIP). In general we consider fields as structures in a language expanding the language of rings, although in the following sentence they are considered in precisely the language of rings. Separably closed fields are stable and bounded PAC fields are simple. There is a considerable amount of work on NIP fields, which include real closed and p-adically closed fields, but this does not concern us in the present paper. We now recall two longstanding open conjectures.

Conjecture 1.1. (1) Infinite stable fields are separably closed.

(2) Infinite simple fields are bounded PAC.

Our general idea is that Conjecture 1.1 is both true and tractable after making the additional assumption of largeness. It is shown in [Johnson et al. 2020] that a large stable field is separably closed. We describe another proof of this result in Section 4A. Here we consider (2), and prove:

Theorem 1.2. *Suppose that K is large and simple. Then there is a bounded* PAC *field L of the same characteristic as K such that the absolute Galois group of L is isomorphic* (*as a topological group*) *to the absolute Galois group of K.*

Note that Theorem 1.2 implies that K is bounded. We prove this separately.

Theorem 1.3. *If K is large and simple then K is bounded.*

The assumption that K is simple can be replaced by the more general assumption that the field K is definable in some model M of a simple theory. If we also require M to be highly saturated we can take K to be type-definable (over a small set of parameters) in M. The latter will follow from our proofs and references and we will not talk about it again. Theorem 1.3 generalizes the theorem of Chatzidakis that a simple PAC field is bounded, which is proven via quite different methods in [Chatzidakis 1999]. Poizat [1983] proved that an infinite stable bounded field is separably closed. Combining Poizat's result with Theorem 1.3 we get the above mentioned result of [Johnson et al. 2020] that large stable fields are separably closed.

Theorem 1.3 is reasonably sharp. The restriction to separable extensions is necessary. If K is separably closed of infinite imperfection degree and $\mathrm{Char}(K) = p > 0$ then K is large, stable, and has infinitely many extensions of degree p. There is an emerging body of work on a generalization of simplicity known as NSOP_1. All known NSOP_1 fields are PAC. Theorem 1.3 fails over NSOP_1 fields as there are unbounded PAC NSOP_1 fields (equivalently, there are PAC fields that are NSOP_1 but not simple). For example if K is characteristic zero, PAC, and the absolute Galois group of K is a free profinite group on $\aleph_0$ generators then K is unbounded and NSOP_1 [Chernikov and Ramsey 2016, Corollary 6.2].

A profinite group G is *projective* if any continuous surjective homomorphism $H \to G$ with H profinite has a section. Ax [1968] showed that the absolute Galois group of a perfect PAC field is projective, and Jarden [1972, Lemma 2.1] proved this for nonperfect PAC fields.

Theorem 1.4. *If K is large and simple then the absolute Galois group of K is projective.*

Theorem 1.2 follows from Theorem 1.3, Theorem 1.4, and the theorem of Lubotzky and van den Dries that for any field K and projective profinite group G there is a PAC field extension L of K such that the absolute Galois group of L is isomorphic to G (see [Fried and Jarden 2005, Corollary 23.1.2]). An earlier version of this paper proved Theorem 1.4 under the additional assumption that K is perfect. Philip Dittmann showed us how to remove this assumption.

Theorem 1.5. *Suppose that K is perfect, large, and simple. Then the Brauer group of K is trivial. It follows that*

(1) *any finite-dimensional division algebra over K is a field, and*

(2) *any Severi–Brauer K-variety V has a K-point.*

We recall the definition of Severi–Brauer variety. Let K^{alg} be an algebraic closure of K. Given a K-variety V we let $V_{K^{\mathrm{alg}}}$ be the base change $V \times_K \operatorname{Spec} K^{\mathrm{alg}}$ of V to a K^{alg}-variety. A *Severi–Brauer variety* is a K-variety V such that $V_{K^{\mathrm{alg}}}$ is isomorphic (over K^{alg}) to $\dim V$-dimensional projective space. A Severi–Brauer variety is geometrically integral, so (2) is a modest step towards the conjecture that large simple fields are PAC. Theorem 1.5 was proven for supersimple fields in [Pillay et al. 1998]; our proof closely follows that in [Pillay et al. 1998], so we do not recall the definition of the Brauer group. (Supersimple fields are perfect, but large simple fields need not be perfect.) Items (1) and (2) of Theorem 1.5 are well-known consequences of triviality of the Brauer group. We refer to [Poonen 2017, Sections 1.5 and 4.5.1] for the definition of the Brauer group and these facts.

Suppose $\mathrm{Char}(K) \neq 2$, then we say that a *conic* over K is a smooth irreducible projective K-curve of genus 0. One-dimensional Severi–Brauer varieties are exactly conics [Poonen 2017, Example 4.5.8]. Thus Theorem 1.6 generalizes the one-dimensional case of Theorem 1.5.2 to imperfect fields.

Theorem 1.6. *Suppose that K is large and simple*, $\mathrm{Char}(K) \neq 2$, *and C is a conic over K. Then C has a K-point*, *hence* (*by largeness*) *$C(K)$ is infinite.*

Let us mention some other earlier work around the conjectures on stable and simple fields described above. One of the first results on deducing algebraic results from model-theoretic hypotheses was Macintyre's theorem [1971] that infinite fields with ω-stable theory are algebraically closed (generalized to superstable fields in [Cherlin and Shelah 1980]). Macintyre's Galois-theoretic method has been used in many later works including the result on large stable fields [Johnson et al. 2020] mentioned above. Supersimple theories are simple theories in which there are not infinite forking chains of types, whereby any complete type has an ordinal valued dimension called the SU-rank. This gives a so-called "surgical dimension" as in [Pillay and Poizat 1995] from which one deduces that an infinite field with *supersimple* theory is perfect and bounded. So insofar as Conjecture 1.1(2) is restricted to supersimple theories, it remained to prove that supersimple theories are PAC, and some partial results were obtained in [Pillay et al. 1998; Martin-Pizarro and Pillay 2004] for example. A theme of the current paper is that, other than perfection of K, any results on supersimple fields also hold over large simple fields.

The conclusions of Theorems 1.4, 1.5, and 1.6 are properties of PAC fields. Another well-known consequence of a field K being PAC is that the Henselization of any nontrivial valuation on K is separably closed; see [Fried and Jarden 2005, Corollary 11.5.9]. In an earlier draft of this paper we showed that any nontrivial valuation on a large simple field has separably closed Henselization. Dittmann generalized this to Theorem 1.7, which is proven in the Appendix.

Theorem 1.7. *Suppose that K is large and* NSOP_∞. *Then any nontrivial valuation on K has separably closed Henselization. In particular*, *any nontrivial valuation on K has algebraically closed residue field and divisible value group.*

NSOP_∞ is a weakening of simplicity; see the Appendix for a definition and some discussion. NSOP_1 implies NSOP_∞ and essentially every known theory without the strict order property is NSOP_∞. It is natural to ask if Theorem 1.7 holds without the assumption of largeness. In the Appendix we give the following partial generalization.

Theorem 1.8. *If K admits a p-valuation then K is* SOP_∞. *In particular if K is a subfield of a finite extension of $\mathbb{Q}_p$ then K is* SOP_∞.

See Section A2 for the definition of a p-valuation. The proof of Theorem 1.8 uses diophantine work of Anscombe, Dittmann, and Fehm [Anscombe et al. 2020] in place of largeness. The results of [Anscombe et al. 2020] are p-adic analogues of classical results on sums of squares. In Section A2 we give a similar argument using Lagrange's four-square theorem to show that a formally real field is SOP_∞. If K admits a valuation with formally real residue field then K is formally real [Bochnak et al. 1998, Corollary 10.1.9]. Thus if K admits a valuation with formally real residue field then K is SOP_∞.

2. Large fields and definability

2A. ***Algebraic conventions.*** We let K^* be the set of nonzero elements of K and $\mathrm{Char}(K)$ be the characteristic of K. A *K-variety* is a separated, reduced K-scheme of finite type. We let $\dim V$ be the usual algebraic dimension and $V(K)$ be the set of K-points of a K-variety V. We let $\mathbb{A}^n$ be n-dimensional affine space over K (recall that $\mathbb{A}^n(K) = K^n$). We often assume irreducibility of the relevant K-varieties. A K-curve is a one-dimensional K-variety. A *morphism* is a morphism of K-varieties.

2B. ***Largeness.*** Large fields were introduced by Florian Pop. A survey appears in [Pop 2014], which starts by saying that large fields are fields over which (or in which) one can do a lot of "interesting mathematics". So largeness looks like a field-arithmetic tameness notion. The field K is *large* if every irreducible K-curve with a smooth (also called nonsingular) K-point has infinitely many K-points.

Fact 2.1 [Pop 1996]. *The following are equivalent*:

(1) *K is large.*

(2) *K is existentially closed in $K((t))$.*

(3) *If an irreducible K-variety V has a smooth K-point then $V(K)$ is Zariski dense in V.*

Fact 2.2 [Pop 2014, Proposition 2.7]. *An algebraic extension of a large field is large.*

Fact 2.3 allows us to pass to elementary extensions.

Fact 2.3 [Pop 2014, Proposition 2.1]. *Large fields form an elementary class.*

2C. ***Existentially étale sets.*** Let W be a K-variety. The authors of [Johnson et al. 2020] introduced the *étale open topology* on $W(K)$. If K is not large then the étale open topology is always discrete and if K is large then the étale open topology on $W(K)$ is nondiscrete whenever $W(K)$ is infinite. Our original proofs were given in terms of this topology, but at present we mostly avoid the topology and give proofs from scratch. We use properties of certain special existentially definable subsets of $W(K)$. A subset X of $W(K)$ is an *EE* set if there is a K-variety V and an étale morphism $f : V \to W$ such that $X = f(V(K))$. It is shown in [Johnson et al. 2020] that the EE subsets of $W(K)$ form a basis for the étale open topology. (In [Johnson et al. 2020] EE sets are referred to as “étale images”.)

If W is smooth and $V \to W$ is an étale morphism then V is also smooth. At present we are mainly concerned with subsets of $K^n = \mathbb{A}^n(K)$, so we may restrict attention to smooth K-varieties. We quickly recall what we need from this setting. Let V, W be smooth irreducible K-varieties. An *étale morphism* $f : V \to W$ is a morphism such that the differential df_a is an isomorphism $TV_a \to TW_{f(a)}$ for all $a \in V$. In particular if $f : \mathbb{A}^n \to \mathbb{A}^n$ is a morphism then f is étale at $a \in K^n$ if and only if the Jacobian of f at a is invertible. The general notion of an étale morphism between not necessarily smooth varieties is more complicated but is not needed here.

Fact 2.4 [EGA IV$_4$ 1967, Proposition 17.1.3]. *Suppose W_1, W_2, V are smooth K-varieties and $f_i : W_i \to V$ is an étale morphism for $i \in \{1, 2\}$. Let W be the fiber product $W_1 \times_V W_2$ and $f : W \to V$ be the canonical map. Then W is a smooth K-variety and f is étale.*

We have $(W_1 \times_V W_2)(K) = \{(a_1, a_2) \in W_1(K) \times W_2(K) : f_1(a_1) = f_2(a_2)\}$, from which it easily follows that the image of $(W_1 \times_V W_2)(K)$ under f agrees with $f_1(W_1(K)) \cap f_2(W_2(K))$. Corollary 2.5 follows.

Corollary 2.5. *Suppose that W is a smooth K-variety. Then the collection of EE subsets of $W(K)$ is closed under finite intersections.*

Corollary 2.5 holds for an arbitrary K-variety, but we do not need this.

Lemma 2.6. *Suppose that K is large, W is a smooth irreducible K-variety, and X is a nonempty EE subset of $W(K)$. Then X is Zariski-dense in W. In particular any nonempty EE subset of K^n is Zariski dense in K^n.*

The identity morphism $W \to W$ is étale, so Lemma 2.6 generalizes the fact that if K is large and W is a smooth irreducible K-variety with $W(K) \neq \varnothing$ then $W(K)$ is Zariski dense in W.

Proof. Let V be a K-variety and $f : V \to W$ be an étale morphism such that $X = f(V(K))$. Suppose that X is not Zariski dense in V. Then X is contained in a proper closed subvariety Y of W. As W is irreducible we have $\dim Y < \dim W$. Note that $f^{-1}(Y)$ is a closed subvariety of V containing $V(K)$. As f is étale it is finite-to-one, hence $\dim V = \dim W$ and $\dim f^{-1}(Y) = \dim Y < \dim W$. So $f^{-1}(Y)$ is a proper closed subvariety of V containing $V(K)$. This contradicts Fact 2.1. $\square$

Corollary 2.7 follows from Corollary 2.5 and Lemma 2.6.

Corollary 2.7. *Suppose that K is large. Let W be a smooth irreducible K-variety and $X_1, \ldots, X_n$ be EE subsets of $W(K)$ with $\bigcap_{i=1}^k X_i \neq \emptyset$. Then $\bigcap_{i=1}^k X_i$ is Zariski dense in W. In particular, if $X_1, \ldots, X_k$ are EE subsets of K^n with $\bigcap_{i=1}^k X_i \neq \emptyset$ then $\bigcap_{i=1}^k X_i$ is Zariski dense in K^n.*

Fact 2.8 is proven in [Johnson et al. 2020] for arbitrary K-varieties.

Fact 2.8. *Let W be a smooth K-variety, $g : W \to W$ be a K-variety isomorphism, and X be an EE subset of $W(K)$. Then $g(X)$ is also an EE subset of $W(K)$.*

Proof. Let V be a smooth K-variety and $f : V \to W$ be an étale morphism such that $X = f(V(K))$. Note that g is étale as any K-variety isomorphism is étale. So $g \circ f : V \to W$ is étale as a composition of étale morphisms is étale. □

We will apply Corollary 2.9 below.

Corollary 2.9. *Suppose that X is an EE subset of K^n, $a = (a_1, \ldots, a_n) \in K^n$, and $b = (b_1, \ldots, b_n) \in (K^*)^n$. Then*

$$X + a = \{(c_1 + a_1, \ldots, c_n + a_n) : (c_1, \ldots, c_n) \in X\},$$
$$bX = \{(b_1c_1, \ldots, b_nc_n) : (c_1, \ldots, c_n) \in X\}$$

are EE subsets of K^n.

Proof. The morphisms $\mathbb{A}^n \to \mathbb{A}^n$ given by $(x_1, \ldots, x_n) \mapsto (x_1 + a_1, \ldots, x_n + a_n)$ and also by $(x_1, \ldots, x_n) \mapsto (b_1x_1, \ldots, b_nx_n)$ are K-variety isomorphisms. Apply Fact 2.8. □

3. Fields with simple theory

We recall some basic results about fields K whose first-order theory is simple, and then make an additional observation under the assumption of largeness. For simple theories see [Kim and Pillay 1997; Casanovas 2011], and for groups definable in (models of) simple theories, see in addition [Pillay 1998; Pillay et al. 1998]. We recall the relevant portions of this theory.

3A. ***Conventions and basic definitions.*** Our model-theoretic notation is standard. We let L be a first-order language, T be a complete consistent L-theory, and $\overline{M}$ be a highly saturated model of T. For now, $x, y, z, \ldots$ range over finite tuples of variables, $a, b, c, \ldots$ range over finite tuples of parameters from $\overline{M}$, and $A, B, C, \ldots$ range over small subsets of $\overline{M}$. "Definable" means "definable in $\overline{M}$, possibly with parameters". We sometimes identify definable sets with the formulas defining them.

Given an L-formula $\varphi(x, y)$ and a suitable tuple b we say that $\varphi(x, b)$ *divides over* a set A of parameters if $\{\varphi(x, b_i) : i < \omega\}$ is inconsistent for some infinite A-indiscernible sequence $(b_i : i < \omega)$ with $b_0 = b$. A partial type $\Sigma(x)$ divides over A if some formula in Σ divides over A. The theory T is *simple* if for any small set A of parameters and complete type $\Sigma(x)$ there is $A_0 \subseteq A$ such that $|A_0| \leq |T|$ and $\Sigma(x)$ *does not divide* over A_0. Simplicity may also be defined in terms of the combinatorial *tree property*, but we do not need this. It is worth mentioning that simplicity is incompatible with the existence of a definable partial ordering which contains an infinite chain. It follows that real closed fields and nonseparably closed Henselian fields are not simple. Nondividing yields a good notion of independence in simple theories: a is *independent* from B over A if $\mathrm{tp}(a/B, A)$ does not divide over A.

3B. ***Generics in definable groups.*** In this section we summarize [Pillay 1998, Section 3], although we introduce things in a different order and use somewhat different terminology. Suppose that T is simple and G is an infinite group definable over $\varnothing$ in $\overline{M}$. A definable subset X of G is (*left*) *f-generic* if every left translate gX of X does not divide over $\varnothing$ and a complete type $\Sigma(x)$ concentrated on G is (left) f-generic if every formula in $\Sigma(x)$ is left f-generic. Note that if a definable $X \subseteq G$ is f-generic then aX is f-generic for any $a \in G$. Note that in [Pillay 1998] "generic" is used for "f-generic". (The language was changed after some more recent work on groups definable in NIP theories.)

Fact 3.1. *Suppose that T is simple*, *G is an $\varnothing$-definable group in $\overline{M}$*, *A*, *B are small sets of parameters*, *and $a \in G$.*

(1) *Left f-genericity is equivalent to right f-genericity* (*so we just say f-generic*).

(2) *If $X \subseteq G$ is f-generic then X is f-generic in any expansion of $\overline{M}$ by constants.*

(3) $\mathrm{tp}(a/A)$ *is f-generic if whenever $b \in G$ is independent from a over A then the product ba is independent of $A \cup \{b\}$ over $\varnothing$.*

(4) *If $A \subseteq B$ and a is independent from B over A*, *then* $\mathrm{tp}(a/B)$ *is f-generic if and only if* $\mathrm{tp}(a/A)$ *is f-generic.*

(5) *If $b \in B$ then* $\mathrm{tp}(a/A, b)$ *is f-generic if and only if* $\mathrm{tp}(ba/A, b)$ *is f-generic.*

(6) *An A-definable subset X of G is f-generic if and only if it is contained in an f-generic complete type over A.*

Fact 3.2 is immediate from the definitions and Fact 3.1.

Fact 3.2. *If X is a definable subset of G which is not f-generic then we have $\bigcap_{i=1}^{k} g_i X = \varnothing$ for some $g_1, \ldots, g_k \in G$.*

Lemma 3.3. *Suppose that T is simple*, *M is a model of T*, *G is an $\varnothing$-definable group in M*, *H is a subgroup of G with $|G/H| \geq \aleph_0$*, *and X is a definable subset of G such that $X \subseteq aH$ for some $a \in G$. Then X is not f-generic. In particular*, *an infinite index definable subgroup of G is not f-generic.*

Proof. Let $(g_i : i < \omega)$ be a sequence of elements of G which lie in distinct cosets of H. So $g_i X \cap g_j X = \varnothing$ when $i \neq j$. After passing to a highly saturated elementary extension and applying Ramsey and saturation we obtain a sequence $(h_i : i < \omega)$ of elements of G which is indiscernible over the defining parameters of X and satisfies $h_i X \cap h_j X = \varnothing$ when $i \neq j$. So X is not f-generic. □

Lemma 3.4. *Suppose that T is simple, X is a definable subset of G, $\approx$ is a definable equivalence relation on X, and each $\approx$-class is f-generic. Then there are only finitely many $\approx$-classes.*

Proof. Suppose towards a contradiction that there are infinitely many $\approx$-classes. Let c be a finite tuple of parameters over which X and $\approx$ are definable. Then there is an $\approx$-class D with canonical parameter d such that $d \notin \operatorname{acl}(c)$. Let $\varphi(x, d, c)$ be a formula defining D and $(d_i : i < \omega)$ be an infinite sequence of realizations of $\operatorname{tp}(d/c)$ which is indiscernible over c and satisfies $d_0 = d$. Then $((c, d_i) : i < \omega)$ is indiscernible, and the formulas $\varphi(x, d_i, c)$ are pairwise inconsistent, so $\varphi(x, d, c)$ divides over $\varnothing$. This contradicts that $\varphi(x, d, c)$ defines the set D which is an f-generic subset of K^n. □

We now prove Lemma 3.5, which we could not find in the literature.

Lemma 3.5. *Suppose T is simple and G, H are $\varnothing$-definable groups in $\overline{M}$. Fix a small set A of parameters and $(a, b) \in G \times H$. Then* $\operatorname{tp}((a, b)/A)$ *is f-generic in $G \times H$ if and only if the following conditions hold*:

(1) $\operatorname{tp}(a/A)$ *is an f-generic type of G*,

(2) $\operatorname{tp}(b/A)$ *is an f-generic type of H*,

(3) *and a is independent from b over A.*

Proof. The definitions and "forking calculus" easily show that (1), (2), and (3) together imply that $\operatorname{tp}((a, b)/A)$ is f-generic in $G \times H$. The difficulty lies in showing that all f-generic types of $G \times H$ are of this form. We suppose that $\operatorname{tp}((a, b)/A)$ is f-generic in $G \times H$. It follows directly that $\operatorname{tp}(a/A)$ and $\operatorname{tp}(b/A)$ are f-generic types of G and H, respectively. It remains to prove that a is independent from b over A. Suppose that $(c, d) \in G \times H$, $\operatorname{tp}(c/A)$, $\operatorname{tp}(d/A)$ are f-generic in G, H, respectively, and (c, d) is independent from (a, b) over A. By Fact 3.1 ca is independent from db over $\varnothing$. As $\operatorname{tp}((a, b)/A)$ is f-generic in $G \times H$, and (a, b) is independent from (c, d) over A, we see that (ca, db) is independent from A, c, d over $\varnothing$. It follows that a is independent from b over A, c, d, and then that a is independent from b over A. □

3C. ***Generics in definable fields.*** Now suppose K is an infinite field definable (say over $\varnothing$) in $\overline{M} \models T$. Everything we say remains true for K a type-definable field in $\overline{M}$. We have two attached groups, the additive group $(K, +)$ and the multiplicative group $(K^*, \times)$ (recall that $K^* = K \setminus \{0\}$). A definable $X \subseteq K$ is *additively f-generic* if it is f-generic in $(K, +)$ and is *multiplicatively f-generic* if $X \cap K^*$ is an f-generic in $(K^*, \times)$, and we make the analogous definitions for a type concentrated on K. The first two claims of Fact 3.6 are from [Pillay et al. 1998, Proposition 3.1]. Uniqueness of f-generic types in stable fields is [Poizat 2001, Theorem 5.10].

Fact 3.6. *Suppose that T is simple. Let X be a definable subset of K, A be a small set of parameters, and $p = \mathrm{tp}(a/A)$ for some $a \in K$. Then the following are equivalent*:

(1) *X is an additive f-generic.*

(2) *X is a multiplicative f-generic.*

Furthermore the following are equivalent:

(1) *p is an additive f-generic*

(2) *p is a multiplicative f-generic.*

If T is stable then there is a unique additive f-generic type over K.

We let D_n be the group $((K^*)^n, \times)$. Corollary 3.7 is a higher-dimensional version of Fact 3.6. The first claim of Corollary 3.7 follows from Fact 3.6, Lemma 3.5, and induction on n. The second claim follows from the first claim and Fact 3.1.5.

Corollary 3.7. *Suppose that T is simple, A is a small set of parameters, $a = (a_1, \dots, a_n) \in K^n$, and $p(x) = \mathrm{tp}(a/A)$. Then p is an f-generic type of $(K^n, +)$ if and only if p is an f-generic type of D_n. So if $X \subseteq K^n$ is definable, then X is f-generic in $(K^n, +)$ if and only if $X \cap D_n$ is f-generic in D_n.*

Proposition 3.8 is our main tool when dealing with large simple fields.

Proposition 3.8. *Suppose that T is simple and K is large. Let X be a definable subset of K^n which contains a nonempty EE subset. Then X is f-generic for $(K^n, +)$, and is hence f-generic for D_n.*

Thus if T is simple and large then any definable subset of K^n with nonempty interior in the étale open topology is f-generic. If K is perfect, bounded PAC then a definable subset of K^n is f-generic if and only if it has nonempty interior in the étale open topology [Walsberg and Ye 2023].

Proof. Suppose towards a contradiction that X is not f-generic for $(K^n, +)$. By Corollary 3.7, $X \cap D_n$ is not f-generic for D_n. We may suppose that X contains $\bar{0} = (0, \dots, 0)$ as both EE subsets and f-generic subsets of K^n are closed under additive translation (by Corollary 2.9 and definitions). Let $X' = X \cap D_n$. By Corollary 3.7, X' is not f-generic in D_n. By Fact 3.2 there are $g_1, \dots, g_k \in D_n$ such that $\bigcap_{i=1}^k g_i X' = \varnothing$. Then $\bigcap_{i=1}^k g_i X$ is nonempty, as it contains $\bar{0}$, but is contained in $K^n \setminus D^n$ and is hence not Zariski dense in K^n. This contradicts Corollary 2.7. □

Fact 3.9 will be crucial for Theorem 1.5. It is proven in [Pillay et al. 1998].

Fact 3.9. *Suppose that T is simple. Let H be a finite index definable subgroup of $(K^*, \times)$ and H_1, H_2 be cosets of H. Then $H_1 + H_2$ contains K^*, namely every nonzero element of K is of the form $a + b$, where $a \in H_1$ and $b \in H_2$.*

4. Proof of Theorem 1.3

This section is the proof of Theorem 1.3. Our proof follows the strategy of the "Remarque" at the end of [Pillay and Poizat 1995] which outlines another proof, suggested by Chatzidakis, of the main result of that paper (that fields equipped with a certain "surgical dimension" are bounded). Remember that when we say that K is bounded we mean that for every n, K has only finitely many extensions of any given degree inside K^{sep}. We first make a few reductions. Fact 4.1 is well-known, but we include a proof for the sake of completeness.

Fact 4.1. *The following are equivalent*:

(1) *K is bounded.*

(2) *For any n there are only finitely many degree n separable extensions of K up to K-algebra isomorphism.*

Proof. By the primitive element theorem a degree n separable extension L of K is of the form $L = K(\alpha)$, where α is a root of a separable irreducible monic degree n polynomial $p(x) \in K[x]$. So L has at most n distinct conjugates over K in K^{sep}; the fact easily follows. □

We set some notation. Given $a = (a_0, \dots, a_{n-1}) \in K^n$ we let $p_a(x)$ denote the polynomial $x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. We let U be the set of $a \in K^n$ such that p_a is separable and irreducible in $K[x]$. Note that U is definable. Given $p \in K[x]$ we let (p) be the ideal in $K[x]$ generated by p. For each $a \in U$ the field extension $K(\alpha)$ generated over K by a root α of p_a is isomorphic to $K[x]/(p_a)$. For $a, b \in U$, we write $a \approx b$ if $K[x]/(p_a)$ is isomorphic over K to $K[x]/(p_b)$. So K has finitely many separable extensions of degree n if and only if there are only finitely many $\approx$-classes.

Remark 4.2. The equivalence relation $\approx$ on U is definable in K.

Proof. The field $K[x]/(p_a)$ is uniformly interpretable in K (as a varies), as an n-dimensional vector space over K (with basis $1, \alpha, \dots, \alpha^{n-1}$ for α a root of $p_a(x)$ and the appropriate multiplication). Now note that if $a, b \in U$ then $a \approx b$ if and only if p_b has a root in $K[x]/(p_a)$. □

Next we have the main result needed to obtain Theorem 1.3:

Theorem 4.3. *Suppose that $a \in U$ and let D be the $\approx$-class of a. Then there is an EE subset X of K^n such that $a \in X \subseteq D$.*

Equivalently: every $\approx$-class is étale open.

Proof of Theorem 4.3. Fix $a \in U$, and let $\alpha \in K^{\mathrm{sep}}$ be a root of $p_a(x)$. Let $\bar{x} = (x_0, \dots, x_{n-1})$ and $\beta(\bar{x}) = x_0 + \alpha x_1 + \cdots + x_{n-1}\alpha^{n-1}$. Let $\alpha = \alpha_1, \dots, \alpha_n$ be the K-conjugates of α, namely the roots of $p_a(x)$ (which are distinct). We write $\beta_i(\bar{x})$ for $x_0 + x_1\alpha_i + \cdots + x_{n-1}\alpha_i^{n-1}$. So, for $b \in K^n$, $\beta_1(b), \dots, \beta_n(b)$ are the K-conjugates of $\beta(b)$.

Let V be the set of $b = (b_0, b_1, \dots, b_{n-1}) \in K^n$ such that $K(\beta(b)) = K(\alpha)$. Note that $b \in V$ if and only if $\beta(b)$ is a root of $p_c(x)$ for some (in fact unique) $c \in U$ such that $c \approx a$. Note further that $b \in V$ if and only if $1, \beta(b), \dots, \beta(b)^{n-1}$ are linearly independent over K, hence V is a Zariski open subset of K^n.

Let $e_1, \dots, e_n \in \mathbb{Z}[\bar{x}]$ be the elementary symmetric polynomials in n variables, i.e.,

$$e_k(\bar{x}) = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} x_{i_1} \cdots x_{i_k}.$$

Given $b = (b_0, \dots, b_{n-1}) \in K^n$ we let

$$G(b) = \big(-e_1(\beta_1(b), \dots, \beta_n(b)), e_2(\beta_1(b), \dots, \beta_n(b)), \\ \dots, (-1)^n e_n(\beta_1(b), \dots, \beta_n(b))\big).$$

Claim 4.4. *There are $G_1, \dots, G_n \in K[\bar{x}]$ such that $G(b) = (G_1(b), \dots, G_n(b))$ for all $b \in K^n$, and if $b \in V$ then $G(b) \approx a$.*

The first claim of Claim 4.4 follows as G is symmetric in $\alpha_1, \dots, \alpha_n$. The second claim follows as $p_{G(b)}$ is the monic polynomial with roots $\beta_1(b), \dots, \beta_n(b)$. Claim 4.5 below is crucial.

Claim 4.5. $G(0, 1, 0, \dots, 0) = a$ *and the Jacobian of* G *at* $(0, 1, 0, \dots, 0)$ *is invertible.*

Given a polynomial function $f : K^n \to K^n$ we let $\mathrm{Jac}_f(a)$ be the Jacobian of f and $|\mathrm{Jac}_f(a)|$ be the Jacobian determinant of f at $a \in K^n$.

Proof. It is clear that $G(0, 1, 0, \dots, 0) = a$ and $(0, 1, 0, \dots, 0) \in V$. Let $L = K(\alpha)$. To show that the Jacobian of G at $(0, 1, 0, \dots, 0)$ is invertible we first produce maps $D, E, F : L^n \to L^n$ such that G agrees with the restriction of $D \circ E \circ F$ to V. We define $F : L^n \to L^n$ by

$$F(b_0, \dots, b_{n-1}) = (b_0 + b_1\alpha_1 + \cdots + b_{n-1}\alpha_1^{n-1}, \dots, b_0 + b_1\alpha_n + \cdots + b_{n-1}\alpha_n^{n-1}).$$

$E : L^n \to L^n$ is given by

$$E(b) = (e_1(b), \dots, e_n(b)),$$

and $D : L^n \to L^n$ is given by

$$D(b_0, \dots, b_{n-1}) = (-b_0, b_1, -b_2, \dots, (-1)^n b_{n-1}).$$

So if $b \in V$ then $G(b) = (D \circ E \circ F)(b)$. Note that F and D are linear, so Jac_F and Jac_D are constant. Applying the chain rule we have

$$\begin{aligned}\mathrm{Jac}_G(0, 1, 0, \dots, 0) &= \mathrm{Jac}_D\, \mathrm{Jac}_E(F(0, 1, 0, \dots, 0))\, \mathrm{Jac}_F \\ &= \mathrm{Jac}_D\, \mathrm{Jac}_E(\alpha_1, \dots, \alpha_n)\, \mathrm{Jac}_F\,.\end{aligned}$$

It is clear that $|\mathrm{Jac}_D| \in \{-1, 1\}$. Furthermore, Jac_F is a Vandermonde matrix

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{n-1} \\ 1 & \alpha_3 & \alpha_3^2 & \cdots & \alpha_3^{n-1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & \alpha_n & \alpha_n^2 & \cdots & \alpha_n^{n-1} \end{pmatrix}.$$

So Jac_F is invertible as $\alpha_1, \dots, \alpha_n$ are distinct. Finally, by [Lascoux and Pragacz 2002],

$$|\mathrm{Jac}_E(\alpha_1, \dots, \alpha_n)| = \prod_{1 \le i < j \le n} (\alpha_i - \alpha_j).$$

This is nonzero as the α_i are distinct, so $\mathrm{Jac}_E(\alpha_1, \dots, \alpha_n)$ is invertible. □

We now deduce Theorem 4.3. Let O be the open subvariety of $\mathbb{A}^n$ given by $|\mathrm{Jac}_G(\bar{x})| \ne 0$. So G gives an étale morphism $O \to \mathbb{A}^n$. Then $O(K) \cap V$ is a Zariski open subset of K^n, which is nonempty by Claim 4.4. Let W be an open subvariety of $\mathbb{A}^n$ such that $W(K) = O(K) \cap V$. The restriction of G to W is an étale morphism $W \to \mathbb{A}^n$. Let $X = G(W(K))$. So X is a nonempty EE subset of K^n contained in the $\approx$-class of a. As a was an arbitrary member of U, this concludes the proof of Theorem 4.3. □

Finally we can complete the proof of Theorem 1.3.

Proof. Let T be a simple theory, M be a model of T, and K be an infinite field definable in M. As remarked at the beginning of this section, it suffices to fix n and show that K has only finitely many separable extensions of degree n, and thus that the definable equivalence relation $\approx$ on the definable set $U \subset K^n$ has only finitely many classes. After possibly passing to an elementary extension we may suppose that M is highly saturated. By Theorem 4.3 and Proposition 3.8, every $\approx$-class is f-generic for $(K^n, +)$. By Lemma 3.4 there are only finitely many $\approx$-classes. □

4A. ***Another proof that large stable fields are separably closed.*** We give a proof that large stable fields are separably closed that avoids Macintyre's Galois-theoretic argument. We first prove Lemma 4.6. We continue to use the notation of the previous section.

Lemma 4.6. *Let Y be the set of $a \in K^n$ such that p_a has n distinct roots in K. Then Y is an EE subset of K^n.*

Proof. Let V be the open subvariety of $\mathbb{A}^n$ given by $x_i \neq x_j$ for all $1 \leq i < j \leq n$. Let $H : K^n \to K^n$ be given by $H(b) = (-e_1(b), e_2(b), \dots, (-1)^n e_n(b))$. So $p_{H(a)}$ is the polynomial with roots $a_1, \dots, a_n$ for any $a = (a_1, \dots, a_n) \in V(K)$. It follows from [Lascoux and Pragacz 2002] that $|\mathrm{Jac}_H(a)|$ agrees up to sign with $\prod_{1 \leq i < j \leq n}(a_i - a_j)$ for any $a = (a_1, \dots, a_n) \in K^n$. So $\mathrm{Jac}_H(a)$ is invertible for all $a \in V(K)$. Thus $H(V(K))$ is an EE subset of K^n. □

We now show that a large stable field is separably closed.

Proof. Suppose that K is large and not separably closed. Fact 3.6 and Lemma 3.5 together show that if K is stable then for each $n \geq 1$ there is a unique n-ary type over K which is generic for $(K^n, +)$. It follows by Proposition 3.8 that if K is stable then any two nonempty EE subsets of K^n have nonempty intersection. As K is not separably closed there is a separable, irreducible, and nonconstant $p \in K[x]$. Suppose that p is monic and fix $a \in K^n$ such that $p = p_a$. By Theorem 4.3 there is an EE subset X of K^n such that $a \in X$ and p_b is separable and irreducible for any $b \in X$. Let Y be the set of $b \in K^n$ such that p_b has n distinct roots in K; by Lemma 4.6, Y is an EE subset of K^n. So X, Y are disjoint nonempty EE subsets of K^n, hence K is unstable. □

The proof above easily adapts to show that an infinite superstable field is algebraically closed. We describe this proof, assuming some familiarity with superstability. We let $\dim_U Z$ be the U-rank of a definable set Z. Suppose that K is infinite and superstable. A superstable field is perfect, so it suffices to show that K is separably closed. Suppose otherwise and fix n such that there is a nonconstant separable irreducible $p \in K[x]$. Let X, Y be as in the proof above. Note that both X and Y contain a set of the form $f(W(K))$, where W is a dense open subvariety of $\mathbb{A}^n$ and $f : W \to \mathbb{A}^n$ is étale. So $\dim_U W(K) = \dim_U K^n$ and the induced map $W(K) \to K^n$ has finite fibers as f is étale. Hence $\dim_U X = \dim_U K^n = \dim_U Y$. So X, Y are both f-generic in $(K^n, +)$, which contradicts uniqueness of generic types.

4B. *Topological corollaries.* Suppose that v is a nontrivial Henselian valuation on K. It follows from the classical Krasner's lemma that each $\approx$-class is open in the v-adic topology on K^n. See for example [Poonen 2017, 3.5.13.2] for a treatment of the case when K is a local field, which easily generalizes to the Henselian case. It is shown in [Johnson et al. 2020] that if K is not separably closed then the v-adic topology on each K^n agrees with the étale open topology. So Corollary 4.7 generalizes this consequence of Krasner's lemma.

Corollary 4.7. *Fix $a \in K^n$ such that p_a is separable and irreducible. Then the set of $b \in K^n$ such that $K[x]/(p_b)$ is K-algebra isomorphic to $K[x]/(p_a)$ is an* *étale open neighborhood of a. So the set of $a \in K^n$ such that p_a is separable and irreducible is étale open.*

Fact 4.8 is proven in [Johnson et al. 2020] by an application of Macintyre's Galois-theoretical argument.

Fact 4.8. *If K is not separably closed then the étale open topology on K is Hausdorff.*

If K is separably closed then the étale open topology agrees with the Zariski topology on $V(K)$ for any K-variety V; equivalently, every EE subset of $V(K)$ is Zariski open. We give a proof of Fact 4.8 which avoids Galois theory. We apply the fact that if $V \to W$ is a morphism between K-varieties then the induced map $V(K) \to W(K)$ is étale open continuous.

Proof. Equip K with the étale open topology. Any affine transformation $x \mapsto ax + b$, $a \in K^*$, $b \in K$ gives a homeomorphism $K \to K$. Thus it is enough to produce two disjoint nonempty étale open subsets of K. The argument of Section 4A yields two disjoint nonempty étale open subsets X, Y of K^n. Fix $p \in X$ and $q \in Y$ and let $f : K \to K^n$ be given by $f(t) = (1-t)p + tq$. Then f is a continuous map between étale open topologies so $f^{-1}(X)$, $f^{-1}(Y)$ are disjoint nonempty étale open subsets of K. $\square$

Finally, we characterize bounded PAC fields amongst PAC fields.

Corollary 4.9. *Suppose that K is* PAC *and equip each K^n with the étale open topology. Then K is bounded if and only if any definable equivalence relation on K^n has only finitely many classes with interior.*

Note that Corollary 4.9 fails when "PAC" is replaced by "large". For example $\mathbb{Q}_p$ is bounded, the étale open topology on $\mathbb{Q}_p$ agrees with the p-adic topology, and the equivalence relation E where $E(a, b)$ if and only if $a, b \in \mathbb{Q}_p$ have the same p-adic valuation is definable and has infinitely many open classes.

Proof. Suppose that K is not bounded. Fix n such that K has infinitely many separable extensions of degree n. Let U and $\approx$ be as in the proof of Theorem 1.3. Then each $\approx$-class is open and there are infinitely many $\approx$-classes. Now suppose that K is bounded and E is a definable equivalence relation on K^n. Note that K is simple. By Proposition 3.8 any E-class with interior is f-generic. The proof of Lemma 3.4 shows that there are only finitely many f-generic E-classes. $\square$

5. Additional remarks and results

We discuss a few related topics and results, and prove Theorem 1.4. If $\mathrm{Char}(K) = p > 0$ then we let $\wp : K \to K$ be the *Artin–Schreier map* $\wp(x) = x^p - x$. This map is an additive homomorphism, so $\wp(K)$ is a subgroup of $(K, +)$. In this section we let $P_n = \{a^n : a \in K^*\}$ for each n. Some of our proofs below could be simplified by applying Scanlon's theorem [Kaplan et al. 2011] that an infinite stable field is Artin–Schreier closed, but we avoid this.

5A. ***Boundedness and large stable fields.*** It is a theorem of Poizat that an infinite bounded stable field is separably closed. Poizat's result and Theorem 1.3 together show that large stable fields are separably closed. Poizat's result is mentioned somewhat informally at the bottom of p. 347 in [Poizat 1983] and does not appear to be well-known, so we take the opportunity to clarify the matter. Fact 5.1 is [Poizat 1983, Lemma 4].

Fact 5.1. *Suppose that L is a finite Galois extension of K. Then the following hold.*

(1) *If $q \neq \mathrm{Char}(K)$ is a prime then there are only finitely many cosets H of P_q in $(K^*, \times)$ such that some (equivalently, any) $a \in H$ has a q-th root in L.*

(2) *Suppose $\mathrm{Char}(K) = p > 0$. Then there are only finitely many cosets H of $\wp(K)$ in $(K, +)$ such that some (equivalently, any) $a \in H$ is of the form $b^p - b$ for some $b \in L$.*

Fact 5.2 follows from Fact 5.1.

Fact 5.2. *Suppose that K is bounded. Then*

(1) *if $q \neq \mathrm{Char}(K)$ is prime then P_q has finite index in $(K^*, \times)$, and*

(2) *if $\mathrm{Char}(K) > 0$ then $\wp(K)$ has finite index in $(K, +)$.*

We sketch a proof. See [Fehm and Jahnke 2016, Lemma 2.2] for a proof of the characteristic zero case of Fact 5.2(1) via Galois cohomology.

Proof. We only prove (1) as the proof of (2) is similar. Suppose $a \in K^*$ and $\alpha \in K^{\mathrm{sep}}$ satisfies $\alpha^q = a$. Then α and its conjugates generate a degree $\leq q$ Galois extension of K. As K is bounded there are only finitely many such extensions. So by Fact 5.1 P_q has finite index in $(K^*, \times)$. □

Finally, Fact 5.3 is essentially proven in [Macintyre 1971] via a Galois-theoretic argument.

Fact 5.3. *Suppose that the following hold for any finite Galois extension L of K:*

(1) *the q-th power map $L^* \to L^*$ is surjective for any prime $q \neq \mathrm{Char}(K)$, and*

(2) *if $\mathrm{Char}(K) \neq 0$ then the Artin–Schreier map $L \to L$ is surjective.*

Then K is separably closed.

We now sketch a proof of Poizat's theorem.

Corollary 5.4. *Suppose that K is infinite, bounded, and stable. Then K is separably closed.*

Proof. We verify the conditions of Fact 5.3. Suppose that L is a finite Galois extension of K. Then L is bounded and stable (the latter holds as L is interpretable in K). As L is stable there is a unique additive (multiplicative) generic type over K (see Fact 3.6). It follows that there are no proper finite index definable subgroups of $(L^*, \times)$ or $(L, +)$. So by Fact 5.2 the q-th power map $L^* \to L^*$ is surjective for any prime $q \neq \mathrm{Char}(K)$ and if $\mathrm{Char}(K) > 0$ then the Artin–Schreier map $L \to L$ is surjective. □

We repeat that the below corollary follows from Fact 5.1 and Theorem 1.3.

Corollary 5.5. *Suppose that K is large and simple. Then*

(1) *if $q \neq \mathrm{Char}(K)$ is prime then P_q has finite index in $(K^*, \times)$, and*

(2) *if $\mathrm{Char}(K) > 0$ then $\wp(K)$ has finite index in $(K, +)$.*

Corollary 5.5(2) is proven more generally for infinite simple fields in [Kaplan et al. 2011]. We take the opportunity to sketch a direct proof of Corollary 5.5. We let $\mathbb{G}_m$ be the scheme-theoretic multiplicative group Spec $K[x, x^{-1}]$, so $\mathbb{G}_m(K) = K^*$.

Proof. We first fix a prime $q \neq \mathrm{Char}(K)$. The morphism $\mathbb{G}_m \to \mathbb{A}^1$ given by $x \mapsto ax^q$ is étale for any $a \in K^*$. So any coset of P_q is an EE subset of K. By the special (and easier) case of Proposition 3.8 when $n = 1$, any coset of P_q is f-generic in $(K^*, \times)$. By Lemma 3.3, P_q has finite index in $(K^*, \times)$. Item (2) follows by a similar argument and the fact that the Artin–Schreier morphism $\mathbb{A}^1 \to \mathbb{A}^1$ is étale. □

Fehm and Jahnke construct an unbounded PAC field K such that the group of n-th powers has finite index in each finite extension of K [Fehm and Jahnke 2016, Proposition 4.4], so Theorem 1.3 does not follow from Corollary 5.5.

5B. ***Conics, Brauer group, and projectivity.*** Corollary 5.6 follows from Fact 3.9 and Corollary 5.5.

Corollary 5.6. *Suppose that K is large and simple, $a, b \in K^*$, and $p \neq \mathrm{Char}(K)$ is a prime. Then there are $c, d \in K$ such that $c^p + ad^p = b$.*

The proof in [Pillay et al. 1998] that conics over (infinite) supersimple fields have points now extends to proving Theorem 1.6.

Proof of Theorem 1.6. Let C be a conic, i.e., a smooth projective irreducible K-curve of genus 0. As $\mathrm{Char}(K) \neq 2$ we may assume that C is a closed subvariety of $\mathbb{P}^2$ given by the homogenous equation $ax^2 + by^2 = z$ for some $a, b \in K^*$. By Corollary 5.6 there are $c, d \in K$ such that $ac^2 + bd^2 = 1$. So $C(K)$ is nonempty. □

We let Br K be the Brauer group of K. Recall that the Brauer group of an arbitrary field is an abelian torsion group. Given a prime p we let $\mathrm{Br}_p\, K$ be the p-part of the Brauer group of K. Facts 5.7 and 5.8 both follow by the proof of [Pillay et al. 1998, Theorem 4.6].

Fact 5.7. *Let $p \neq \mathrm{Char}(K)$ be a prime. Suppose that whenever L is a finite separable extension of K and $a \in L^*$, then $\{b^p + ac^p : b, c \in L^*\}$ contains L^*. Then $\mathrm{Br}_p K$ is trivial.*

Fact 5.8. *Suppose that*

(1) *K is perfect, and*

(2) *if L is a finite extension of K, p is a prime, and $a \in L^*$, then $\{b^p + ac^p : b, c \in L^*\}$ contains L^*.*

Then the Brauer group of K is trivial.

We now prove Theorem 1.5.

Proof. It suffices to show that the second condition of Fact 5.8 is satisfied. Let L be a finite extension of K and p be a prime. Note that L is perfect as a finite extension of a perfect field is perfect; the case when $p = \mathrm{Char}(K)$ follows. Suppose that $p \neq \mathrm{Char}(K)$. Note that L is simple as L is interpretable in K and L is large by Fact 2.2. Apply Corollary 5.6. □

Theorem 1.4 follows from Proposition 5.9 as a field of cohomological dimension ≤ 1 has projective absolute Galois group [Gruenberg 1967].

Proposition 5.9. *If K is simple and large then K has cohomological dimension ≤ 1.*

See [Serre 1997, Chapter I, §3] for an overview of cohomological dimension. The proof of Proposition 5.9 is due to Philip Dittmann. The simpler case where K is assumed to be perfect was proved earlier by the authors. We do not know if every large simple field has trivial Brauer group.

Proof. Suppose K is simple and large. The same argument as in the proof of Theorem 1.5 shows that if $p \neq \mathrm{Char}(K)$ is a prime, L is a finite separable extension of K, and $a \in L^*$, then $\{b^p + ac^p : b, c \in L^*\}$ contains L^*. So by Fact 5.7, $\mathrm{Br}_p L$ is trivial for every finite extension L of K and prime $p \neq \mathrm{Char}(K)$. By [Serre 1997, II.2.3 Proposition 4], K has p-cohomological dimension ≤ 1 for every prime $p \neq \mathrm{Char}(K)$. By [Serre 1997, II.2.2 Proposition 3], any field L has $\mathrm{Char}(L)$-cohomological dimension ≤ 1. So K has cohomological dimension ≤ 1. □

Appendix: NSOP$_\infty$ fields

by Philip Dittmann

A theory T has the *fully finite strong order property* if there is a formula $\psi(x, y)$, with the two tuples of variables x and y having equal length, a model $M \models T$, and a sequence $(a_i)_{i\in\omega}$ of tuples in M satisfying $M \models \psi(a_i, a_j)$ for all $i < j$, and for any $n \geq 3$ the formula $\psi(x_1, x_2) \wedge \cdots \wedge \psi(x_{n-1}, x_n) \wedge \psi(x_n, x_1)$ is inconsistent with T. In short, the binary relation described by ψ admits infinite chains and does not admit cycles. In this situation we also say that T has or is SOP_∞. A structure M is SOP_∞ if its theory is. A theory or structure is NSOP_∞ if it is not SOP_∞.

In Shelah's terminology, T having the fully finite strong order property witnessed by ψ is equivalent to ψ having the n-strong order property SOP_n for T, for all $n \geq 3$ [Shelah 1996, Definition 2.5]. In particular, if T is complete and simple then T is NSOP_∞ [Shelah 1996, Claim 2.7]. The notion "fully finite strong order property" seems to have first appeared in an unpublished manuscript by Adler [2008], although it has by now also been used elsewhere [Conant and Terry 2016, Definition 2.1].

A1. ***Valuations.*** The Henselization of a PAC field with respect to any nontrivial valuation is separably closed [Fried and Jarden 2005, Corollary 11.5.9]. Thus the following can be seen as supporting evidence for the conjecture that large simple fields are PAC.

Theorem A.1. *Suppose that K is large and v is a nontrivial valuation on K. If (K, v) has nonseparably closed Henselization then K is* SOP_∞. *In particular, if either the residue field of v is not algebraically closed or the value group of v is not divisible then K is* SOP_∞.

The second claim of Theorem A.1 follows from the first as the Henselization of (K, v) has the same residue field and value group as (K, v) [Engler and Prestel 2005, Theorem 5.2.5], and a nontrivially valued separably closed field has algebraically closed residue field and divisible value group [Engler and Prestel 2005, Theorem 3.2.11]. We will make use of Fact A.2, proven in [Johnson et al. 2020, Theorem 6.15].

Fact A.2. *Let v be a nontrivial valuation on K. If the Henselization of (K, v) is not separably closed then the étale open topology refines the v-adic topology on K.*

The following argument using generics was used in a preliminary version of the main article to get the simple case of Theorem A.1. Suppose that K is simple and the Henselization of (K, v) is not separably closed. By Fact A.2, $\mathfrak{m}_v$ is an étale open neighborhood of 0, so there is an EE subset U of K satisfying $0 \in U \subset \mathfrak{m}_v$. By Proposition 3.8 the set U is f-generic for $(K, +)$. This contradicts Lemma 3.3 as $\mathfrak{m}_v$ is an infinite index subgroup of $(K, +)$. This argument does not generalize to large NSOP_1 fields as at present there is no theory of generics in NSOP_1 groups. (This is not straightforward: [Dobrowolski 2020] gives an example of a definable group in an NSOP_1 structure in which generics with respect to Kim forking do not exist.)

Lemma A.3. *Suppose that K is large and $U \subseteq K$ is an étale open neighborhood of zero. Then for any $n \geq 2$ there is $a \in K^*$ such that $a, a^2, \ldots, a^n \in U$.*

Proof. For each $i \in \{2, \ldots, n\}$ let $V_i = \{b \in K : b^i \in U\}$. Each map $K \to K, b \mapsto b^i$ is continuous with respect to the étale open topology, so each V_i is an étale open neighborhood of zero. Then $V = V_1 \cap \cdots \cap V_n$ is an étale open neighborhood of zero. As K is large V contains a nonzero element of K. □

Proof of Theorem A.1. By Fact A.2 there is a nonempty EE subset U of K with $0 \in U \subseteq \mathfrak{m}_v$. Let $\psi(x, y)$ be the formula $(x \neq 0) \wedge (y \neq 0) \wedge (x^{-1}y \in U)$. Note that if $K \models \psi(a, b)$ then $b/a \in \mathfrak{m}_v$, hence $v(a) < v(b)$. We show that $\psi(x, y)$ witnesses SOP_∞. First suppose that $a_1, \ldots, a_n \in K$ satisfy

$$\psi(a_1, a_2) \wedge \cdots \wedge \psi(a_{n-1}, a_n) \wedge \psi(a_n, a_1).$$

Then we have $v(a_1) < v(a_2) < \cdots < v(a_{n-1}) < v(a_n) < v(a_1)$, a contradiction. We now show that for each $n \geq 1$ there are $a_1, \ldots, a_n \in K$ such that $K \models \psi(a_i, a_j)$ if and only if $i < j$. By Lemma A.3 there is $a \in K^*$ such that $a, a^2, \ldots, a^n \in U$. Then $K \models \psi(a^i, a^j)$ for $i < j$. Thus the binary relation on K defined by ψ admits chains of arbitrary finite length, and in a saturated elementary extension of K we obtain an infinite chain. Thus K is SOP_∞. □

Recall that EE sets are existentially definable. Note that the formula ψ in the proof of Theorem A.1 is existential. This is optimal as a quantifier-free formula in an arbitrary field is stable. This is similar to the result, proven in [Johnson et al. 2020, Theorem 3.1], that an unstable large field admits an unstable existential formula. The witnesses for SOP_∞ produced in the next section are also existential.

If v is actually Henselian, the same technique as in the proof of Theorem A.1 gives a slightly stronger statement. This is presumably well-known to the experts, but appears not to be available in the literature.

Theorem A.4. *Suppose that v is a nontrivial Henselian valuation on K and K is not separably closed. Then K has the strict order property* [Shelah 1996, Definition 2.1].

Proof. Fact A.2 provides an EE subset U of K with $0 \in U \subseteq \mathfrak{m}_v$. By [Johnson et al. 2020, Theorem B] the v-topology on K agrees with the étale open topology, hence U is v-open, and in particular contains a ball around 0.[1] Therefore, for any element $c \in K^\times$ with $v(c)$ sufficiently large we have $cU \subsetneq U$. Thus the definable family $\{xU : x \in K\}$ contains the infinite chain $U \supsetneq cU \supsetneq c^2U \supsetneq \cdots$ under inclusion. Hence K has the strict order property. □

A2. ***Formally real and formally p-adic fields.*** Corollary 5.6 implies that if K is large, simple, and of characteristic zero then there are $a, b \in K$ such that $a^2+b^2 = -1$.

[1]This argument does not seriously use the étale open topology — we only need that the topology given by v is definable in the field language. This latter fact is already implicit in [Prestel and Ziegler 1978, Remark 7.11].

So a large simple field cannot be formally real. Duret [1977] showed that formally real fields are unstable. Theorem A.5 generalizes these.

Theorem A.5. *Suppose that K is formally real. Then K is* SOP_∞.

Proof. Let $\varphi(x, y)$ be the formula

$$\exists z_1, z_2, z_3, z_4\, [x - y - 1 = z_1^2 + z_2^2 + z_3^2 + z_4^2].$$

We show that φ witnesses SOP_∞. An application of Lagrange's four-square theorem shows that $K \models \varphi(m, m')$ for all integers $m > m'$. Now suppose that $a_1, \dots, a_n \in K$ and we have $K \models \{\varphi(a_1, a_2), \dots, \varphi(a_{n-1}, a_n), \varphi(a_n, a_1)\}$. Then

$$-n = (a_1 - a_2 - 1) + (a_2 - a_3 - 1) + \cdots + (a_{n-1} - a_n - 1) + (a_n - a_1 - 1)$$

is a sum of squares, a contradiction. □

Fix a prime p. A field K is p-adically closed if K is elementarily equivalent to a finite extension of $\mathbb{Q}_p$ and K is *formally p-adic* if K embeds into a p-adically closed field. An equivalent definition (which we shall not need) is that there exists a p-valuation v on K, i.e., v is of mixed characteristic, the residue field is a finite extension of $\mathbb{F}_p$, and the interval $[0, v(p)]$ in the value group is finite. Indeed, if v is a p-valuation on K then the so-called p-adic closure of (K, v) is an elementary extension of a finite extension of $\mathbb{Q}_p$. See [Prestel and Roquette 1984] for a comprehensive treatment of formally p-adic fields.

Theorem A.6. *Suppose that K is formally p-adic. Then K is* SOP_∞.

Proof. Let F be a finite extension of $\mathbb{Q}_p$ such that K embeds into an elementary extension of F. Let v be the unique extension of the p-adic valuation on $\mathbb{Q}_p$ to F and $\mathcal{O}_F$ be the valuation ring of v.

By [Anscombe et al. 2020, Propositions 4.7 and 4.8] (applied to the base field $K = \mathbb{Q}$, the prime $\mathfrak{p} = p$ of $\mathbb{Q}$, and the relative type τ of $F/\mathbb{Q}_p$ in the terminology there), there exists a parameter-free existential formula $\psi(x)$ such that $\psi(F) \subseteq \mathcal{O}_F$, and $\psi(\mathbb{Q}) = \mathbb{Z}_{(p)}$. (Note that the paper cited phrases the result in terms of a concrete "diophantine family" $D^{\tau}_{\mathfrak{p}, t_{\mathfrak{p}}, A, B}$, but this is effectively the same as an existential formula with parameters from the base field $\mathbb{Q}$ [Anscombe et al. 2020, Remark 3.2], and parameters from $\mathbb{Q}$ can be eliminated.)

Let $\varphi(x, y)$ be the formula

$$(y \neq 0) \wedge \exists z\, (\psi(z) \wedge y = p \cdot x \cdot z).$$

We show that $\varphi(x, y)$ witnesses SOP_∞ for K. Suppose $m < m'$ are integers. Then we have $\mathbb{Q} \models \varphi(p^m, p^{m'})$, since $p^{m'}/(p \cdot p^m) \in \mathbb{Z} \subseteq \psi(\mathbb{Q})$. Since φ is existential, we have $K \models \varphi(p^m, p^{m'})$. Thus the binary relation on K defined by φ admits an infinite chain.

Now suppose that K satisfies

$$\Theta = \exists x_1, \dots, x_n\, [\varphi(x_1, x_2) \wedge \cdots \wedge \varphi(x_{n-1}, x_n) \wedge \varphi(x_n, x_1)].$$

As Θ is existential and K embeds into an elementary extension of F, we have $F \models \Theta$. Hence there are $b_1, \dots, b_n \in F$ such that

$$F \models \{\varphi(b_1, b_2), \dots, \varphi(b_{n-1}, b_n), \varphi(b_n, b_1)\}.$$

As $\psi(F) \subseteq \mathcal{O}_F$, we see that $F \models \varphi(a, a')$ implies that $v(a) < v(a')$ for any $a, a' \in F$. Thus we have $v(b_1) < v(b_2) < \cdots < v(b_{n-1}) < v(b_n) < v(b_1)$, a contradiction. $\square$

Acknowledgements

Pillay was supported by NSF grants DMS-1665035, DMS-1760212, and DMS-2054271. We thank Daniel Max Hoffmann for finding some mistakes, and the referee for suggestions.

References

[Adler 2008] H. Adler, "Around the strong order property and mock simplicity", unpublished manuscript, 2008.

[Anscombe et al. 2020] S. Anscombe, P. Dittmann, and A. Fehm, "A p-adic analogue of Siegel's theorem on sums of squares", *Math. Nachr.* **293**:8 (2020), 1434–1451. MR Zbl

[Ax 1968] J. Ax, "The elementary theory of finite fields", *Ann. of Math.* (2) **88** (1968), 239–271. MR Zbl

[Bochnak et al. 1998] J. Bochnak, M. Coste, and M.-F. Roy, *Real algebraic geometry*, Ergebnisse der Math. (3) **36**, Springer, 1998. MR Zbl

[Casanovas 2011] E. Casanovas, *Simple theories and hyperimaginaries*, Lect. Notes in Logic **39**, Association for Symbolic Logic, Chicago, 2011. MR Zbl

[Chatzidakis 1999] Z. Chatzidakis, "Simplicity and independence for pseudo-algebraically closed fields", pp. 41–61 in *Models and computability* (Leeds, UK, 1997), edited by S. B. Cooper and J. K. Truss, London Math. Soc. Lecture Note Ser. **259**, Cambridge Univ. Press, 1999. MR Zbl

[Cherlin and Shelah 1980] G. Cherlin and S. Shelah, "Superstable fields and groups", *Ann. Math. Logic* **18**:3 (1980), 227–270. MR Zbl

[Chernikov and Ramsey 2016] A. Chernikov and N. Ramsey, "On model-theoretic tree properties", *J. Math. Log.* **16**:2 (2016), art. id. 1650009. MR Zbl

[Conant and Terry 2016] G. Conant and C. Terry, "Model theoretic properties of the Urysohn sphere", *Ann. Pure Appl. Logic* **167**:1 (2016), 49–72. MR Zbl

[Dobrowolski 2020] J. Dobrowolski, "Sets, groups, and fields definable in vector spaces with a bilinear form", preprint, 2020. arXiv 2004.07238

[Duret 1977] J.-L. Duret, "Instabilité des corps formellement réels", *Canad. Math. Bull.* **20**:3 (1977), 385–387. MR Zbl

[EGA IV$_4$ 1967] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas, IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR Zbl

[Engler and Prestel 2005] A. J. Engler and A. Prestel, *Valued fields*, Springer, 2005. MR Zbl

[Fehm and Jahnke 2016] A. Fehm and F. Jahnke, "Fields with almost small absolute Galois group", *Israel J. Math.* **214**:1 (2016), 193–207. MR Zbl

[Fried and Jarden 2005] M. D. Fried and M. Jarden, *Field arithmetic*, 2nd ed., Ergebnisse der Math. (3) **11**, Springer, 2005. MR Zbl

[Gruenberg 1967] K. W. Gruenberg, "Projective profinite groups", *J. London Math. Soc.* **42** (1967), 155–165. MR Zbl

[Jarden 1972] M. Jarden, "Elementary statements over large algebraic fields", *Trans. Amer. Math. Soc.* **164** (1972), 67–91. MR Zbl

[Johnson et al. 2020] W. Johnson, C.-M. Tran, E. Walsberg, and J. Ye, "The étale-open topology and the stable fields conjecture", preprint, 2020. arXiv 2009.02319

[Junker and Koenigsmann 2010] M. Junker and J. Koenigsmann, "Schlanke Körper (slim fields)", *J. Symbolic Logic* **75**:2 (2010), 481–500. MR Zbl

[Kaplan et al. 2011] I. Kaplan, T. Scanlon, and F. O. Wagner, "Artin–Schreier extensions in NIP and simple fields", *Israel J. Math.* **185** (2011), 141–153. 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

[Lascoux and Pragacz 2002] A. Lascoux and P. Pragacz, "Jacobians of symmetric polynomials", *Ann. Comb.* **6**:2 (2002), 169–172. MR Zbl

[Macintyre 1971] A. Macintyre, "On ω_1-categorical theories of fields", *Fund. Math.* **71**:1 (1971), 1–25. MR Zbl

[Martin-Pizarro and Pillay 2004] A. Martin-Pizarro and A. Pillay, "Elliptic and hyperelliptic curves over supersimple fields", *J. London Math. Soc.* (2) **69**:1 (2004), 1–13. MR Zbl

[Pillay 1998] A. Pillay, "Definability and definable groups in simple theories", *J. Symbolic Logic* **63**:3 (1998), 788–796. MR Zbl

[Pillay and Poizat 1995] A. Pillay and B. Poizat, "Corps et chirurgie", *J. Symbolic Logic* **60**:2 (1995), 528–533. MR Zbl

[Pillay et al. 1998] A. Pillay, T. Scanlon, and F. O. Wagner, "Supersimple fields and division rings", *Math. Res. Lett.* **5**:4 (1998), 473–483. MR Zbl

[Poizat 1983] B. Poizat, "Groupes stables, avec types génériques réguliers", *J. Symbolic Logic* **48**:2 (1983), 339–355. MR Zbl

[Poizat 2001] B. Poizat, *Stable groups*, Math. Surveys and Monographs **87**, Amer. Math. Soc., Providence, RI, 2001. MR Zbl

[Poonen 2017] B. Poonen, *Rational points on varieties*, Grad. Studies in Math. **186**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Pop 1996] F. Pop, "Embedding problems over large fields", *Ann. of Math.* (2) **144**:1 (1996), 1–34. MR Zbl

[Pop 2014] F. Pop, "Little survey on large fields — old & new", pp. 432–463 in *Valuation theory in interaction* (Segovia and El Escorial, Spain, 2011), edited by A. Campillo et al., Eur. Math. Soc., Zürich, 2014. MR Zbl

[Prestel and Roquette 1984] A. Prestel and P. Roquette, *Formally p-adic fields*, Lect. Notes in Math. **1050**, Springer, 1984. MR Zbl

[Prestel and Ziegler 1978] A. Prestel and M. Ziegler, “Model-theoretic methods in the theory of topological fields”, *J. Reine Angew. Math.* **299-300** (1978), 318–341. MR Zbl

[Serre 1997] J.-P. Serre, *Galois cohomology*, Springer, 1997. MR Zbl

[Shelah 1996] S. Shelah, “Toward classifying unstable theories”, *Ann. Pure Appl. Logic* **80**:3 (1996), 229–255. MR Zbl

[Walsberg and Ye 2023] E. Walsberg and J. Ye, “Éz fields”, *J. Algebra* **614** (2023), 611–649. MR Zbl

Received 4 May 2022. Revised 2 Jan 2023.

ANAND PILLAY:

anand.pillay.3@nd.edu
Department of Mathematics, University of Notre Dame, Notre Dame, IN, United States

ERIK WALSBERG:

Department of Mathematics, University of California, Irvine, CA, United States

PHILIP DITTMANN:

philip.dittmann@tu-dresden.de
Institut für Algebra, Technische Universität Dresden, Dresden, Germany

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.381

Additive reducts of real closed fields and strongly bounded structures

Hind Abu Saleh and Ya'acov Peterzil

Given a real closed field R, we identify exactly four proper reducts of R which expand the underlying (unordered) R-vector space structure. Towards this theorem we introduce the new notion of strongly bounded reducts of linearly ordered structures: a reduct $\mathcal{M}$ of a linearly ordered structure $\langle R; <, \dots \rangle$ is called *strongly bounded* if every $\mathcal{M}$-definable subset of R is either bounded or cobounded in R. We investigate strongly bounded additive reducts of o-minimal structures and prove the above theorem on additive reducts of real closed fields.

1. Introduction

The study of *ordered* additive reducts of real closed fields starts with the work of Pillay, Scowcroft and Steinhorn [Pillay et al. 1989], followed by Marker, Peterzil and Pillay [Marker et al. 1992]. The motivation behind the work here is a conjecture about *unordered* such reducts from [Peterzil 1993]. Before stating the conjecture, let us clarify our usage of the notion of "reduct" here.

Definition 1.1. Given two structures $\mathcal{M}$ and $\mathcal{N}$, we say that $\mathcal{M}$ is a *reduct of* $\mathcal{N}$ (or, $\mathcal{N}$ is an *expansion of* $\mathcal{M}$), denoted by $\mathcal{M} \dot{\subseteq} \mathcal{N}$, if $\mathcal{M}$ and $\mathcal{N}$ have the same universe and every set that is definable in $\mathcal{M}$ is also definable in $\mathcal{N}$ (where definability allows parameters). We say that $\mathcal{M}$ and $\mathcal{N}$ are *interdefinable*, denoted by $\mathcal{M} \doteq \mathcal{N}$, if $\mathcal{M}$ is a reduct of $\mathcal{N}$ and $\mathcal{N}$ is reduct of $\mathcal{M}$.

We say $\mathcal{M}$ is *a proper reduct* of $\mathcal{N}$ (or $\mathcal{N}$ a *proper expansion* of $\mathcal{M}$) if $\mathcal{M} \dot{\subseteq} \mathcal{N}$ and not $\mathcal{M} \doteq \mathcal{N}$.

Below, we let Λ_R be the family of all R-linear maps $\lambda_a(x) = \alpha x$ for all $\alpha \in R$. Our ultimate goal here is to prove the following:

Peterzil is partially supported by the Israel Science Foundation grant 290/19.
MSC2020: 03C64.
Keywords: additive reducts of real closed fields, strongly bounded structures.

Theorem 1.2. *Let R be a real closed field. Then the only reducts between the vector space $\langle R; +, \Lambda_R\rangle$ and the field $\langle R; <, +, \cdot\rangle$ are as follows*:

$$\mathcal{R}_{\mathrm{alg}} = \langle R; +, \cdot, <\rangle,$$
$$\mathcal{R}_{\mathrm{sb}} = \langle R; +, <, \Lambda_R, \mathfrak{B}\rangle,$$
$$\mathcal{R}_{\mathrm{semi}} = \langle R; +, <, \Lambda_R\rangle, \qquad \mathcal{R}_{\mathrm{bd}} = \langle R; +, <^*, \Lambda_R, \mathfrak{B}\rangle,$$
$$\mathcal{R}^*_{\mathrm{lin}} = \langle R; +, <^*, \Lambda_R\rangle,$$
$$\mathcal{R}_{\mathrm{lin}} = \langle R; +, \Lambda_R\rangle,$$

where $<^*$ is the linear order on the interval $(0, 1)$ *and $\mathfrak{B}_{\mathrm{sa}}$ the collection of all bounded semialgebraic sets over R.*

Remark 1.3. (1) The definable sets in $\mathcal{R}_{\mathrm{alg}}$ are called semialgebraic, while those definable in $\mathcal{R}_{\mathrm{semi}}$ are semilinear. The structure $\mathcal{R}_{\mathrm{sb}}$ above is called *semibounded*, as it expands the ordered vector space by a collection of bounded sets. Semibounded structures were studied in several articles, for example, [Edmundo 2000; Belegradek 2004; Peterzil 2009].

(2) Notice that because all the above structures expand the full underlying R-vector space, then once $<^*$ is definable then the restriction of $<$ to every bounded interval is definable.

(3) A similar project, in the setting of Presburger arithmetic, was carried out in [Conant 2018], where it was proven that there are no proper reducts between $\langle \mathbb{Z}; +\rangle$ and $\langle \mathbb{Z}; <, +\rangle$. We expect that in arbitrary models of Presburger arithmetic, an analogous result to Theorem 1.2 holds, with the intermediate reducts corresponding to possible restrictions of $<$ to infinite subintervals.

Some of the work towards the proof of Theorem 1.2 can be read off earlier results. In particular, the fact that the semibounded reduct $\mathcal{R}_{\mathrm{sb}}$ is the only proper reduct between $\mathcal{R}_{\mathrm{semi}}$ and $\mathcal{R}_{\mathrm{alg}}$ was proven over $\mathbb{R}$ in [Peterzil 1993] and can be deduced for arbitrary real closed fields from [Edmundo 2000] (see Fact 5.1 below). However, the bulk of the work here is to show that if a reduct $\mathcal{M}$ of $\mathcal{R}_{\mathrm{alg}}$ does not define the full order then it is necessarily a reduct of $\mathcal{R}_{\mathrm{bd}}$. Towards that, we introduce a new notion of "a strongly bounded structure" in a more general setting, and most of our results here are about such structures.

Definition 1.4. Let $\mathcal{R} = \langle R; <, \dots\rangle$ be a linearly ordered structure. A reduct $\mathcal{M} = \langle R; \dots\rangle$ of $\mathcal{R}$ is called *strongly bounded* if every $\mathcal{M}$-definable $X \subseteq R$ is either bounded or cobounded (namely, $R \setminus X$ is bounded).

Remark 1.5. (1) The term "strongly bounded" was chosen to reflect a combination of a semibounded structure with a strongly minimal one. Almost all of our work here concerns strongly bounded additive reducts of o-minimal structures, where the underlying linear order is dense. Analogous definitions could be given for, say, models of Presburger arithmetic if one wishes to study all reducts which expand the underlying ordered group.

(2) The definition of a strongly bounded structure requires an ambient linear order. Thus it might not seem amenable to working in elementarily equivalent structures. However, in practice we only work in sufficiently saturated elementary extensions of a strongly bounded $\mathcal{M}$ as above, and thus we may assume that this elementary extension is also a reduct of a linearly ordered elementary extension of $\mathcal{R}$.

By definition, if $\mathcal{M}$ is a strongly bounded reduct of a linearly ordered structure then the ordering $<$ is not definable in $\mathcal{M}$. We prove several results about strongly bounded reducts of o-minimal structures (see, for example, Theorems 4.5 and 4.27):

Theorem. *Let $\langle R; <, +, \dots \rangle$ be an o-minimal expansion of an ordered group and let $\mathcal{M} = \langle R, +, \dots \rangle$ be a strongly bounded reduct.*

(1) *Every $\mathcal{M}$-definable subset of R^n is already definable in $\langle R; +, \Lambda_{\mathcal{M}}, \mathfrak{B}^* \rangle$, where Λ_M is the collection of $\mathcal{M}$-definable endomorphisms of $\langle R, + \rangle$ and $\mathfrak{B}^*$ is the collection of all $\mathcal{M}$-definable bounded sets.*

(2) *For every $\mathcal{N} \equiv \mathcal{M}$, the model theoretic algebraic closure equals the definable closure.*

2. Proper expansions of $\mathcal{R}_{\mathrm{lin}}$

In this section we assume that $\mathcal{R}_{\mathrm{omin}}$ is an o-minimal expansion of a real closed field R and $\mathcal{M} = \langle R; +, \dots \rangle$ is an additive reduct of $\mathcal{R}_{\mathrm{omin}}$.

Theorem 2.1. *If $\mathcal{M}$ is not a reduct of $\mathcal{R}_{\mathrm{lin}} = \langle R; +, \Lambda_R \rangle$, then $<^*$ is definable in $\mathcal{M}$.*

Proof. It is sufficient to prove that some interval $[0, b]$ is $\mathcal{M}$-definable, for $b > 0$.

Claim 2.2. $\mathrm{Th}(\mathcal{M})$ *is unstable.*

Proof. This is based on work of Hasson, Onshuus and Peterzil [Hasson et al. 2010].

Assume towards contradiction that $\mathrm{Th}(\mathcal{M})$ is stable. By [Hasson et al. 2010, Theorem 1], every 1-dimensional stable structure interpretable in an o-minimal structure is necessarily 1-based. So $\mathcal{M}$ is 1-based. By [Hrushovski and Pillay 1987, Theorem 4.1], it follows that every $\mathcal{M}$-definable set is a boolean combination of cosets of definable subgroups of R^n. Every definable subgroup of $\langle R^n; + \rangle$ in an o-minimal structure is an R-vector subspace of R^n and therefore every $\mathcal{M}$-definable set is definable in $\mathcal{R}_{\mathrm{lin}}$, a contradiction. Hence $\mathcal{M}$ is unstable. □

Because $\mathcal{M}$ is unstable, it is in particular not strongly minimal. This generally implies that in some elementary extension of $\mathcal{M}$, we have an $\mathcal{M}$-definable subset in one variable which is infinite and coinfinite. However, o-minimal structures eliminate $\exists^\infty$, and therefore so does $\mathcal{M}$. It follows that there is some $\mathcal{M}$-definable subset of R itself which is infinite and coinfinite. Call this set Y.

In the special case where both Y and $R \setminus Y$ are unbounded in R we can prove a stronger result which will be used several times here, and thus we state it separately.

Lemma 2.3. *Assume that $Y \subseteq R$ is definable in an o-minimal expansion of an ordered group. If both Y and $R \setminus Y$ are unbounded then the full linear order is definable in $\langle R; +, Y\rangle$.*

Proof. By o-minimality, Y has the form

$$Y := I_1 \cup I_2 \cup \cdots \cup I_n \cup L, \tag{1}$$

such that for every $i \in \{1, \ldots, n\}$, $I_i := (a_i, b_i)$, L is a finite set and in addition $-\infty \le a_1 < b_1 < a_2 < \cdots < a_n < b_n \le +\infty$. Without loss of generality $L = \varnothing$.

Since both Y and $R \setminus Y$ are unbounded, Y has the form (1) above and without loss of generality, we may assume that $I_1 = (-\infty, b_1)$, and $I_i = (a_i, b_i)$ for $i \in \{2, \ldots, n\}$.

By replacing Y by $Y - b_1$ we may assume that $b_1 = 0$, and then

$$-Y \cap Y = (-b_n, -a_n) \cup \cdots \cup (-b_2, -a_2) \cup (a_2, b_2) \cup \cdots \cup (a_n, b_n).$$

So $(-Y \cap Y) \cap [(-Y \cap Y) + (a_n + b_n)]$ equals the interval $I_n = (a_n, b_n)$ in Y. Replace Y by $Y_1 := Y \setminus I_n$; now Y_1 contains an unbounded ray together with $n-2$ bounded intervals. Continuing in this way we obtain a ray $(-\infty, 0)$ that is definable, so we can define $<$. $\square$

In the remaining case, either Y or $R \setminus Y$ are bounded, so we assume that Y is bounded, and as above

$$Y := (a_1, b_1) \cup \cdots \cup (a_, b_n),$$

with $a_i, b_i \in M$.

Let $\alpha := b_n - b_1$. The set $(Y + \alpha) \cap Y$ defines a single interval whose right endpoint is b_n. So, we are done. If Y is unbounded then replace Y by $R \setminus Y$ and finish as before. Hence, we have showed that $<^*$ is definable in $\mathcal{M}$. $\square$

3. Reducts of $\mathcal{R}_{\text{alg}}$ which are not semilinear

Here R is a real closed field and $\mathcal{R}_{\text{alg}} = \langle R; <, +, \cdot\rangle$. Before the next theorem we recall previous work from [Loveys and Peterzil 1993] (see a corrected and more general proof in [Belegradek 2004]), which will be used in its proof.

Given $a > 0$ in R, let $I = (-a, a)$. Denote by $+^*$ the partial function obtained by intersecting the graph of $+$ with I^3, and for each $\alpha \in R$, let λ^*_α be the partial function obtained by intersecting the graph of λ_α with I^2. Finally, let $<^*$ be the restriction of $<$ to I^2. Notice that for each $X \subseteq R^n$ such that $\langle R; <, +, \cdot, X\rangle$ is o-minimal, the structure

$$\mathcal{I} = \langle \mathrm{I}; <^*, +^*, \{\lambda^*_\alpha\}_{\alpha\in R}, X \cap I^n\rangle$$

is o-minimal as well.

In [Loveys and Peterzil 1993] the structure $\langle I; <^*, +^*\rangle$ was called a group-interval and its o-minimal expansions were studied there.

A partial endomorphism (p.e. for short) of this group-interval was a function $f : I \to I$ which respects addition when defined: namely, if $x, y, x +^* y \in I$ then $f(x +^* y) = f(x) +^* f(y)$.

Notice that in our setting every $\mathcal{I}$-definable p.e. is necessarily the restriction of λ_α for some $\alpha \in R$. Indeed, if $f : I \to I$ is an $\mathcal{I}$-definable p.e. then it is not hard to verify that

$$H = \{r \in R : \exists\varepsilon > 0\, \forall x \in (-\varepsilon, \varepsilon)\ f(rx) = rf(x)\}$$

is a semialgebraic subgroup of $\langle R, +\rangle$ which contains all integers.

O-minimality of the real field implies that $H = R$ and therefore f is the restriction of an R-linear map, namely the restriction of λ_α for some R.

Now, without going through their precise definition of “a linear theory”, it was shown in [Loveys and Peterzil 1993, Proposition 4.2] that if $\mathrm{Th}(\mathcal{I})$ is linear then every $\mathcal{I}$-definable set is already defined in the structure $\langle I; +^*, <^*, \{\lambda^*_\alpha\}_{\alpha\in R}\rangle$ (possibly together with additional parameters). Thus if $\mathrm{Th}(\mathcal{I})$ is linear then $X \cap I^n$ is a semilinear set.

The following proposition seems to be obvious but for the sake of completion we include a proof in the Appendix.

Fact 3.1. *Let R be a real closed field and $X \subseteq R^n$ a definable set in an o-minimal expansion of $\langle R; <, +, \cdot\rangle$. If X is not semilinear then, in $\mathcal{M} = \langle R; <^*, +, \Lambda_R, X\rangle$, there exists a definable bounded set which is not semilinear.*

Theorem 3.2. *If $X \subseteq R^n$ is semialgebraic and not definable in $\mathcal{R}_{\mathrm{semi}}$, then every bounded R-semialgebraic set is definable in $\langle R; +, \Lambda_R, X\rangle$.*

Proof. Let $\mathcal{M} := \langle R; +, \Lambda_R, X\rangle$. By Theorem 2.1, the relation $<^*$ is definable in $\mathcal{M}$. Let us first see that $\mathcal{M}$ defines a real closed field on some interval.

By Fact 3.1, we may assume that $X \cap I^n$ is not semilinear, for some bounded interval $I = (-a, a)$. Consider the o-minimal structure

$$\mathcal{I} := \langle I; <^*, +^*, \{\lambda^*_\alpha\}_{\alpha\in R}, X \cap I^n\rangle,$$

as we described before stating the theorem. We noted that if $\mathrm{Th}(\mathcal{I})$ is linear then the set $X \cap I^n$ must be semilinear. Because $X \cap I^n$ is not semilinear then $\mathrm{Th}(\mathcal{I})$ is not linear in the sense of [Loveys and Peterzil 1993]. Therefore, by [Peterzil and

Starchenko 1998, Theorem 1.2], a real closed field is $\mathcal{I}$-definable, and hence also $\mathcal{M}$-definable, on some interval $J \subseteq I$.

Without loss of generality, assume that $J = (-a_0, a_0)$, $a_0 > 0$. Denote the field by

$$\mathcal{J} = \langle J, \oplus, \odot \rangle.$$

The structure $\mathcal{J}$ is $\mathcal{M}$-definable. By [Peterzil 1993, Corollary 2.4], every R-semialgebraic subset of J^k, $k \in \mathbb{N}$, is definable in $\mathcal{J}$, and therefore in $\mathcal{M}$.

Let $B \subseteq (-b, b)^n$ for some $b > 0$ in R. Using scalar multiplication from Λ_R, we can contract $(-b, b)$ into $(-a_0, a_0)$, so it is definable in $\mathcal{J}$. It follows that B is definable in $\mathcal{M}$. □

4. Strongly bounded structures

The ultimate goal of this section is to prove:

Theorem 4.1. *Let R be a real closed field. If $X \subseteq R^n$ is semialgebraic and not definable in $\mathcal{R}_{\text{bd}} = \langle R; <^*, +, \Lambda_R, \mathfrak{B}_{\text{sa}} \rangle$, then $<$ is definable in $\langle R; +, \Lambda_R, X \rangle$.*

We are going to work in a more general setting than that of a real closed field. Recall that a strongly bounded reduct of a linearly ordered $\langle R; <, \dots \rangle$ is one in which every definable subset of R is bounded or cobounded. Below, we will mostly be interested in strongly bounded reducts of o-minimal structures. By Lemma 2.3 and the definition of a strongly bounded structure, we have:

Lemma 4.2. *Let $\mathcal{R}_{\text{omin}} = \langle R; <, +, \dots \rangle$ be an o-minimal expansion of an ordered group. If $\mathcal{M} = \langle R; +, \dots \rangle$ is a reduct of $\mathcal{R}_{\text{omin}}$ then $\mathcal{M}$ is strongly bounded if and only if $<$ is not definable in $\mathcal{M}$.*

So in order to prove Theorem 4.1 it is sufficient to prove that if $X \subseteq R^n$ is definable in a strongly bounded $\mathcal{M} = \langle R; +, \dots \rangle$ then X is definable in $\langle R; +, \Lambda_{\mathcal{M}}, \mathfrak{B}_{\mathcal{M}} \rangle$, where $\mathfrak{B}_{\mathcal{M}}$ is the collection of all $\mathcal{M}$-definable bounded sets. A more precise and slightly stronger theorem — Theorem 4.5 — will be proved soon. We first make a general observation which we shall exploit repeatedly.

Definability of "boundedness". For $X \subseteq T \times R^n$, $T \subseteq R^m$ and $t \in T$, we let

$$X_t = \{a \in R^n : \langle t, a \rangle \in X\}.$$

The following general result will be very useful here.

Proposition 4.3. *Let $\mathcal{M} = \langle R; +, \dots \rangle$ be any reduct of an o-minimal expansion of an ordered group. If $\{X_t : t \in T\}$ is an $\mathcal{M}$-definable family of subsets of R^n, then the set*

$$\{t \in T : X_t \text{ is bounded in } R^n\}$$

is definable in $\mathcal{M}$.

Proof. Note that a set $Y \subseteq R^n$ is bounded if and only if for each i, the image of Y under the projection map $\pi_i : \langle y_1, \ldots, y_n \rangle \mapsto y_i$ is bounded in R. Thus, it is sufficient to prove the result under the assumption that all X_t are subsets of R.

By o-minimality, each $X_t \subseteq R$ is unbounded if and only if it contains an unbounded ray. Thus, it is easy to see that

$$\{t \in T : X_t \text{ is bounded}\} = \{t \in T : \exists a\ (a + X_t \cap X_t = \varnothing)\},$$

and hence the set is definable in $\mathcal{M}$. □

The strongly bounded setting. We first clarify and somewhat generalize our setting.

Let $\mathcal{R}_{\mathrm{omin}} = \langle R, <, +, \ldots \rangle$ denote an o-minimal expansion of an ordered group in the language $\mathcal{L}_{\mathrm{omin}}$, and let $\mathcal{M} = \langle R; +, \ldots \rangle$ denote a strongly bounded reduct of $\mathcal{R}_{\mathrm{omin}}$, in the language $\mathcal{L}$, such that $\mathrm{acl}_{\mathcal{M}}(\varnothing)$ *contains at least one nonzero element* (*it follows that* $\mathrm{acl}_{\mathcal{M}}(\varnothing)$ *is infinite*).

Definition 4.4. An interval $(a, b) \subseteq R$ is called a *$\varnothing$-interval in $\mathcal{M}$* if $a, b \in \mathrm{acl}_{\mathcal{M}}(\varnothing)$. A subset $X \subseteq R^n$ is called *$\varnothing$-bounded in $\mathcal{M}$* if X is contained in some I^n, for I a $\varnothing$-interval in $\mathcal{M}$.

Our standing assumption is that for every $\varnothing$-interval $I \subseteq R$, the restricted order $<\restriction I$ is $\varnothing$-definable in $\mathcal{M}$. Notice that, using Theorem 2.1, this is true when $\mathcal{M}$ is elementarily equivalent to a reduct of a real closed field which properly expands $\mathcal{R}_{\mathrm{lin}}$.

We let $\Lambda_{\mathcal{M}}$ be the collection of all $\mathcal{M}$-definable endomorphisms of $\langle R, + \rangle$, defined over $\varnothing$. We let $\mathcal{L}_{\mathrm{bd}}(\mathcal{M})$ be the language consisting of $\{+, \{\lambda\}_{\lambda \in \Lambda_{\mathcal{M}}}\}$, augmented by a predicate for every $\varnothing$-definable, $\varnothing$-bounded set in $\mathcal{M}$.

By expanding $\mathcal{L}$ and $\mathcal{L}_{\mathrm{omin}}$ by function symbols and predicates for $\varnothing$-definable sets, we may assume that

$$\mathcal{L}_{\mathrm{bd}} \subseteq \mathcal{L} \subseteq \mathcal{L}_{\mathrm{omin}}.$$

We let $\mathcal{M}_{\mathrm{bd}}$ be the reduct of $\mathcal{M}$ to $\mathcal{L}_{\mathrm{bd}}$.

Our ultimate goal in this section is to prove:

Theorem 4.5. *For $\mathcal{M}$ strongly bounded as above, every definable subset of R^n is definable in $\mathcal{M}_{\mathrm{bd}}$.*

One of our main difficulties in working with strongly bounded structures is the failure of global cell decomposition. For instance, the set $R \setminus \{0\}$ cannot be decomposed definably into definable cells in a strongly bounded structure, because no ray is definable there.

Another difficulty is the fact that a priori we do not know whether the model theoretic algebraic closure equals the definable closure in strongly bounded structures. However, we shall eventually show in Theorem 4.27 that acl = dcl in this setting.

We assume for the rest of this section that $\mathcal{M}$ is strongly bounded as above.

Definable subsets of R in strongly bounded structures. Notice that although the full order is not definable in $\mathcal{M}$, a basis for the $<$-topology on R and the product topology on R^n is definable in $\mathcal{M}$, using the restricted order. Thus we have:

Lemma 4.6. *If $\{X_t : t \in T\}$ is an $\mathcal{M}$-definable family of subsets of R^n, then the families*

$$\{\mathrm{Cl}(X_t) : t \in T\}, \qquad \{\mathrm{Int}(X_t) : t \in T\}, \qquad \{\mathrm{Fr}(X_t) : t \in T\}$$

are definable in $\mathcal{M}$.

Every $\mathcal{M}$-definable $X \subseteq R$ is a union of finitely many pairwise disjoint *maximal* open subintervals of X (which are possibly not $\mathcal{M}$-definable) and a finite set. Below, when we say that I *is an interval in* X we mean that I is one of these open components of X.

Definition 4.7. Let $Y \subseteq R$ be an $\mathcal{M}$-definable set. We define

$$\partial^-(Y) := \{y \in R : y \text{ is a left endpoint of an interval in } Y\},$$
$$\partial^+(Y) := \{y \in R : y \text{ is a right endpoint of an interval in } Y\}.$$

Lemma 4.8. *If $\{Y_t : t \in T\}$ is an $\mathcal{M}$-definable family of bounded subsets of R, then the families $\{\partial^-(Y_t) : t \in T\}$, $\{\partial^+(Y_t) : t \in T\}$ are $\mathcal{M}$-definable over the same parameter set.*

Proof. We fix an $\mathcal{M}$-definable $<\restriction(0, a_0)$ for some $a_0 > 0$. We define $\partial^-(Y_t)$ by the formula

$$(x \notin Y_t \wedge \exists \varepsilon < a_0\, (x, x+\varepsilon) \subseteq Y_t)$$
$$\vee \big(x \in Y_t \wedge \exists \varepsilon \le a_0\, (x-\varepsilon, x) \cap Y_t = \varnothing \wedge (x, x+\varepsilon) \subseteq Y_t\big).$$

Because of the definability of $<^*$ in $\mathcal{M}$, $\{\partial^-(Y_t) : t \in T\}$ is $\mathcal{M}$-definable. We similarly handle $\partial^+(Y_t)$. $\square$

The next theorem is an important component of our analysis of strongly bounded structures.

Theorem 4.9. *If $\{X_t : t \in T\}$ is an $\mathcal{M}$-definable family of bounded subsets of R, then there is a uniform bound on the length of each interval in X_t. Moreover, there exists such a bound in* $\mathrm{dcl}_{\mathcal{M}}(\varnothing)$.

Proof. By Proposition 4.3, every $\mathcal{M}$-definable family $\{X_t : t \in T\}$ of bounded subsets of R is a subfamily of a $\varnothing$-definable family of such sets. Namely, if $\varphi(x, t, a)$ is the formula defining the X_t's over a, as t varies, then we can consider the formula

$$\psi(x, t, y) : \varphi(x, t, y) \wedge\ \psi(R, t, y) \text{ is a bounded set.}$$

Thus, it is sufficient to prove the result for $\varnothing$-definable families.

By Lemma 4.6, we may assume that each X_t is an open set. We use induction on the maximum number n of intervals in X_t, for $t \in T$.

For $n = 1$, write $X_t = (a_t, b_t)$. Consider the family $\{X_t - a_t : t \in T\}$. By Lemma 4.8, the family is $\varnothing$-definable. Thus, the set $Y = \bigcup_{t \in T} X_t - a_t$ is an $\mathcal{M}$-definable interval, over $\varnothing$, whose left endpoint is 0. Because $\mathcal{M}$ is strongly bounded, this interval must be bounded, and hence its right endpoint is some $K \in M$. By Lemma 4.8, the point K is definable over $\varnothing$.

Consider now the case $n = k+1$, i.e., each X_t consists of at most $k+1$ pairwise disjoint open intervals. For each $t \in T$, let $D_t = \{c_1 - c_2 : c_1, c_2 \in \partial^-(X_t)\}$, an $\mathcal{M}$-definable set by Lemma 4.8.

Claim 4.10. *For each $t \in T$, there exists $d \in D_t$ such that $(X_t + d) \cap X_t$ is one of the intervals in X_t.*

Proof. Let $X_t = I_{1,t} \cup I_{2,t} \cup \cdots \cup I_{k+1,t}$, where each $I_{m,t} := (a_{m,t}, b_{m,t})$, such that

$$a_{1,t} < b_{1,t} < a_{2,t} < b_{2,t} < \cdots < a_{k+1,t} < b_{k+1,t}.$$

For an interval $I = (a, b)$, let $|I| = b - a$.

Let $d = a_{k+1,t} - a_{1,t}$. In the set $X_t + d$, for each m, the interval $I_{m,t}$ is shifted to $I_{m,t} + d$. So $(X_t + d) \cap X_t$ consists of either $I_{k+1,t}$ (when $|I_{k+1,t}| < |I_{1,t}|$) or $I_{1,t} + d$ (when $|I_{k+1,t}| > |I_{1,t}|$).

If it consists of I_{k+1} we are done. Otherwise we take

$$d' = a_{1,t} - a_{k+1,t} \in D_t$$

and then $(X_t + d') \cap X_t = I_{1,t}$.

So in both cases there exists $d \in D_t$ such that $X_t + d \cap X_t$ is one of the intervals in X_t. □

We define the set

$$D'_t := \{d \in D_t : (X_t + d) \cap X_t \text{ is one of the intervals in } X_t\}.$$

Claim 4.11. *The family $\{D'_t : t \in T\}$ is an $\mathcal{M}$-definable family of nonempty sets.*

Proof. For $t \in T$, $d \in D'_t$ if and only if the following two statements hold:

(1) $\partial^-((X_t + d) \cap X_t) \subseteq \partial^-(X_t)$ and $|\partial^-((X_t + d) \cap X_t)| = 1$, and

(2) $\partial^+((X_t + d) \cap X_t) \subseteq \partial^+(X_t)$ and $|\partial^+((X_t + d) \cap X_t)| = 1$.

By Lemma 4.8, (1) and (2) are definable properties in $\mathcal{M}$. By Claim 4.10, each D'_t is nonempty. □

We proceed with the proof of Theorem 4.9. Consider the $\mathcal{M}$-definable family

$$\{Y_{t,d} := X_t + d \cap X_t : d \in D'_t,\ t \in T\},$$

still defined in $\mathcal{M}$ over $\varnothing$. For every t and $d \in D'_t$, the set $Y_{t,d}$ consists of a single interval which is one of the intervals in X_t. By case $n = 1$ we know that there is a uniform bound w_1 on the length of each $Y_{t,d}$, which can be chosen in $\mathrm{dcl}_{\mathcal{M}}(\varnothing)$. We now define, still over $\varnothing$, the family

$$\{Z_{t,d} := X_t \setminus Y_{t,d} : d \in D'_t,\ t \in T\}.$$

Each subset $Z_{t,d}$ consists of at most k intervals among the $k+1$ intervals of X_t. By the induction hypothesis, we know that there is a uniform bound w_2 on the length of each interval, which we may choose in $\mathrm{dcl}_{\mathcal{M}}(\varnothing)$.

Thus the maximum of w_1, w_2, which is in $\mathrm{dcl}_{\mathcal{M}}(\varnothing)$, is the bound on the length of each interval of X_t, as t varies. This ends the proof of Theorem 4.9. $\square$

As a corollary we can now match, definably in $\mathcal{M}$, each left endpoint of an interval in X_t with the corresponding right endpoint:

Proposition 4.12. *Let $\{X_t : t \in T\}$ be an $\mathcal{M}$-definable family of bounded subsets of R, and let*

$$L_t = \{\langle a, b\rangle \in \partial^-(X_t) \times \partial^+(X_t) : \text{the interval } (a, b) \text{ is one the intervals of } X_t\}.$$

Then the family $\{L_t : t \in T\}$ is $\mathcal{M}$-definable.

Proof. By Theorem 4.9, there is a bound $K \in \mathrm{dcl}_{\mathcal{M}}(\varnothing)$ for the length of each interval in X_t, for all $t \in T$. For each $t \in T$, we have

$$\langle a, b\rangle \in L_t \iff a \in \partial^-(X_t) \quad \text{and} \quad b = \min(\partial^+(X_t) \cap [a, a+K]). \qquad (*)$$

By Lemma 4.8, $\partial^-(X_t)$ and $\partial^+(X_t)$ are definable families and since in $(*)$ we only use the order on $[0, K]$, the family $\{L_t : t \in T\}$ is definable in $\mathcal{M}$. $\square$

Remark 4.13. (1) Notice that Theorem 4.9 fails without the assumption that the X_t's are bounded sets. Namely, it is not true in general that the lengths of the bounded components of X_t are bounded in t. For example, the set $X_t = R \setminus \{-t, t\}$ has $(-t, t)$ as an open component, with unbounded length as $t \to \infty$.

Also, even if each X_t is bounded it is not true that the diameter of the X_t's is uniformly bounded. For example, take the family $\{(-t, t-1) \cup (t, t+1) : t \in R\}$ that is definable using $< \restriction (0, 1)$.

(2) We do not know whether Proposition 4.12 holds if we drop the assumption that the X_t's are bounded. Can we still match definably the left and right endpoints of the bounded components of X_t, when the X_t's are unbounded?

Affine sets and functions. Recall that $\mathcal{R}_{\mathrm{omin}}$ is an o-minimal expansion of an ordered divisible abelian group R, and we assume that $\mathcal{M} = \langle R; +, \dots\rangle$ is a strongly bounded reduct of $\mathcal{R}_{\mathrm{omin}}$ in which $<$ is $\varnothing$-definable on every $\varnothing$-interval. We let $<^*$ denote the ordering on some fixed interval we call $(0, 1)$.

Definition 4.14. Let $\langle R; <, +\rangle$ be an abelian ordered divisible group.

(1) A map $f : R^n \to R^k$ is *affine* if it is of the form $\ell(x) + d$ for $\ell : R^n \to R^k$ a homomorphism between $\langle R^n, +\rangle$ and $\langle R^k, +\rangle$, and $d \in R^k$.

(2) A (partial) function $f : R \to R$ is *eventually affine* if there exists $a > 0$ such that $(a, \infty) \subseteq \mathrm{dom}(f)$ and the restriction of f to $(a, +\infty)$ is affine.

(3) $X \subseteq R^n$ is *locally affine at* $a \in X$ if there is an open neighborhood $U \ni a$ such that for all $x, y, z \in U \cap X$, $x - y + z \in X$. The *affine part of* X is the set

$$\mathscr{A}(X) = \{x \in X : X \text{ is locally affine at } x\}.$$

Notice that if X is the graph of an affine map then $\mathscr{A}(X) = X$. Conversely, if X is the graph of a definable function from an open subset of R^k into R^ℓ and $a = (a', f(a')) \in \mathcal{A}(X)$ then f is an affine map in a neighborhood of a_1.

Because a basis for the R^n-topology is definable in $\mathcal{M}$, we immediately have:

Lemma 4.15. *Let $\{X_t : t \in T\}$ be an $\mathcal{M}$-definable family of subsets of R^n, defined over $\varnothing$. Then the family $\{\mathscr{A}(X_t) : t \in T\}$ is $\mathcal{M}$-definable over $\varnothing$.*

Proposition 4.16. *Every $\mathcal{M}$-definable endomorphism $f : R \to R$ is $\varnothing$-definable.*

Proof. Assume that f is defined by an $\mathcal{M}$-formula $\varphi(x, y, a)$ over the parameter a. We show that f can be defined without parameters.

Since being an R-endomorphism is $\mathcal{M}$-definable, we may assume that there is some $\mathcal{M}$-definable $T \subseteq R^k$ such that for all $t \in T$, if $\varphi(R^2, t)$ is nonempty then it defines a nonzero endomorphism f_t of $\langle R; +\rangle$.

Assume first that the set of endomorphisms f_t defined by φ is finite. Define $t_1 E t_2$ if $f_{t_1} = f_{t_2}$, an $\mathcal{M}$-definable equivalence relation. Consider the functions near 0, and define $[t_1]_E < [t_2]_E$ if for all $x > 0$ sufficiently small, we have $f_{t_1}(x) < f_{t_2}(x)$. By o-minimality, we obtain a linear ordering of the finitely many E-classes, and since $<$ is $\mathcal{M}$-definable in a neighborhood of 0, this ordering is $\mathcal{M}$-definable. Thus, each f_t in this finite family of endomorphisms is $\varnothing$-definable.

Assume now that the family $\{f_t : t \in T\}$ is infinite, and we shall reach a contradiction. Consider the set $\{f_t(1) : t \in T\}$. By o-minimality it contains an open interval (a, b), and by replacing each f_t with $f_t - f_{t_0}$, for some $t_0 \in T$ for which $f_{t_0} \in (a, b)$, we may assume that the interval (a, b) contains 0 and the ordering on (a, b) is $\mathcal{M}$-definable (we think of $f_t(a)$ as "the slope" of f_t). Let $T_0 = \{t \in T : f_t(1) \in (0, b)\}$.

We write $t_1 \sim t_2$ if $f_{t_1} = f_{t_2}$, and let $[t]$ be the equivalence class of t. In abuse of notation we let $f_{[t]}$ denote the corresponding endomorphism of R.

By o-minimality, if $f_{t_1}(1) = f_{t_1}(1)$ then $f_{t_1} = f_{t_2}$. Thus we obtain an $\mathcal{M}$-definable function $t : (0, b) \to T_0/\sim$, defined by $f_{[t(x)]}(1) = x$. Namely, $f_{[t(x)]}$ is the endomorphism whose "slope" is x. Fix an element $d > 0$, and define $\sigma : (0, b) \to R$

by $\sigma(x) = f^{-1}_{[t(x)]}(d)$. Namely, $\sigma(x) = y$ if there exists $t \in T_0$ such that $f_t(1) = x$ and $f_t(y) = d$ (we may think of $\sigma(x)$ as "d/x"). The function σ is also $\mathcal{M}$-definable. For every $t \in T_0$, we have $f_t(1) > 0$, and hence $f_t(x) > 0$ if and only if $x > 0$. Therefore, σ is positive on $(0, b)$.

Claim. $\mathrm{Im}(\sigma)$ *is unbounded in* R.

Indeed, assume towards contradiction that $K = \sup(\mathrm{Im}(\sigma)) < \infty$. By our observation, $K > 0$. Choose $y_0 \in \mathrm{Im}(\sigma)$, $y_0 < K$ and sufficiently close to K such that $K < 2y_0$. By assumption, there exists $t_0 \in T_0$ and $x_0 > 0$ such that $f_{t_0}(1) = x_0$ and $f_{t_0}(y_0) = d$.

Let $t_1 \in T_0$ be such that $[t_1] = t(x_0/2)$. Then $f_{t_1}(1) = x_0/2 = f_{t_0}(1)/2$. It follows that $f_{t_1} = f_{t_0}/2$ and hence

$$f_{t_1}(2y_0) = f_{t_0}(2y_0)/2 = f_{t_0}(y_0) = d.$$

But then $f_{t_1}(1) = x_0/2$ and $f_{t_1}(2y_0) = d$, so by definition, $\sigma(x_0/2) = 2y_0 > K$, contradicting the assumption that K bounds $\mathrm{Im}(\sigma)$.

Thus, $\mathrm{Im}(\sigma)$ is an $\mathcal{M}$-definable set which is unbounded and positive, contradicting the assumption that $\mathcal{M}$ is strongly bounded. $\square$

Definition 4.17. We denote by Λ_{omin} the set of all $\mathcal{R}_{\mathrm{omin}}$-definable endomorphisms $f : \langle R, +\rangle \to \langle R, +\rangle$, and we still let $\Lambda_{\mathcal{M}}$ denote the set of all $\mathcal{M}$-definable endomorphisms of R, which by Proposition 4.16, is necessarily $\varnothing$-definable. Let $\Lambda^*_{\mathrm{omin}}$ and $\Lambda^*_{\mathcal{M}}$ denote those nonzero endomorphisms.

Definable functions of one variable. Our goal is to describe definable functions in one variable, and prove that $\mathcal{M}$ has no definable "poles".

Proposition 4.18. *If* $g : R \to R$ *is an* $\mathcal{M}$*-definable partial function whose domain is cobounded and* $\mathrm{Im}(g)$ *is bounded, then* g *is constant on a cobounded set.*

Proof. By o-minimality, there exists $L \in R$ such that $\lim_{x\to+\infty} g(x) = L$. We shall see that $g \equiv L$ on a cobounded set.

The function g is definable in an o-minimal structure, so there exists $a_1 \in R$ such that $g \restriction (a_1, +\infty)$ is either constant or strictly monotone, and there exists a_2 such that g is constant or strictly monotone on $(-\infty, a_2)$.

If g is constant L on $(a_1, +\infty)$ then $\{x \in R : g(x) = L\}$ is unbounded and since $\mathcal{M}$ is strongly bounded the set must be cobounded and we are done. Assume towards contradiction that $g \restriction (a_1, \infty)$ is strictly monotone.

Assume first that g is strictly increasing on (a_1, ∞). Notice that the property of being locally increasing in a neighborhood of $x \in R$ is definable using $<^*$. Thus the set

$$\{x \in R : g \text{ is locally increasing at } x\}$$

is $\mathcal{M}$-definable, contains (a_1, ∞) and hence must be cobounded. It follows that g is strictly increasing on $(-\infty, a_2)$.

Because $\lim_{x\to\infty} g(x) = L$ and g is increasing, there exists $b \in R$ such that for all $x > b$, $L-1 < g(x) < L$. Because $<^*$ is $\mathcal{M}$-definable the set of all $x \in R$ such that $L-1 < g(x) < L$ is $\mathcal{M}$-definable, so must be cobounded. In particular, we may assume that $L-1 < g(x) < L$ for all $x < a_2$ and thus $g(x)$ has a limit $L_1 \in R$ as $x \to -\infty$.

But since g is increasing on $x < a_2$, it follows that $L_1 < L$ and in addition there exists $a_2' \leq a_2$ and $\varepsilon > 0$, such that for all $x < a_2'$,

$$L_1 < g(x) < L_1 + \varepsilon < L.$$

Using $<^*$ again, this is an $\mathcal{M}$-definable property of x so must hold also for all $x > a_1'$, contradicting the fact that $\lim_{x\to+\infty} g(x) = L$.

A similar argument works when g is eventually decreasing. □

Remark 4.19. By [Edmundo 2000], if $\mathcal{N} = \langle R; <, +, \dots \rangle$ is an o-minimal expansion of an ordered group in which every definable bounded function is eventually constant then $\mathcal{N}$ is *semibounded*, namely every definable set is definable using the underlying ordered vector space, together with all the definable bounded sets. This might suggest a fast deduction of Theorem 4.5 from Proposition 4.18. The problem of this approach is that we do not know that the definable functions in the strongly bounded $\mathcal{M} = \langle R; +, <^*, \dots \rangle$ are the same as in its expansion by the full $<$. Thus, we do not see how to apply Edmundo's theorem here.

Next, using almost identical arguments to [Edmundo 2000] we show that every $\mathcal{M}$-definable function $f : R \to R$ is affine on a cobounded set. For that, we recall some notation and facts, based on [Miller and Starchenko 1998].

Notation. For $\mathcal{R}_{\text{omin}}$-definable positive (partial) functions $f, g : R \to R$ such that $(a, \infty) \subseteq \text{dom}(f), \text{dom}(g)$, we write $f \leq g$ (or $f < g$) if $f(x) \leq g(x)$ (or $f(x) < g(x)$) for all large enough x.

We write $v(f) < v(g)$ if $|f| > |\lambda \circ g|$ for all $\lambda \in \Lambda^*_{\text{omin}}$ such that $\lambda > 0$. We also write $v(f) = v(g)$ if there are $\lambda_1, \lambda_2 \in \Lambda^*_{\text{omin}}$, both positive, such that

$$|\lambda_1 \circ g| \leq |f| \leq |\lambda_2 \circ g|.$$

This is easily seen to be an equivalence relation, which roughly says that the rate of growth of f and g at $+\infty$ is of the same scale. In the case where R expands a real closed field then $v(f) = v(g)$ if and only if f and g belong to the same Archimedean class with respect to R, namely there exists $r \in R$ such that $(1/r)|g| \leq f \leq r|g|$.

Finally, we write $\Delta(f) = f(x+1) - f(x)$.

Fact 4.20 [Edmundo 2000]. *For every $\mathcal{R}_{\text{omin}}$-definable function on an unbounded ray*,

(1) *if* $v(f) > v(x)$ *then* $\lim_{x\to\infty} \Delta(f) = 0$;

(2) *if* $v(f) < v(x)$ *then* $v(f^{-1}) > v(x)$;

(3) *if* $v(f) = v(x)$ *then* $\Delta(f)(x)$ *has a limit in* R *as* $x \to \infty$.

The following is just a warm-up towards Theorem 4.25. The proof follows closely the proof of [Edmundo 2000, Proposition 2.8], which uses results of [Miller and Starchenko 1998].

Lemma 4.21. *If* $f : R \to R$ *is* $\mathcal{M}$*-definable on a cobounded set, then* f *is eventually affine. Moreover, there exists a* $\varnothing$*-definable endomorphism* $\lambda \in \Lambda_{\mathcal{M}}$ *and* $A > 0$ *such that for all* x *with* $|x| > A$, *we have* $f(x) = \lambda(x) + d$ *for some* $d \in R$.

Proof. Assume towards contradiction that $f : R \to R$ is not eventually affine. Without loss of generality, f is eventually increasing, and by Proposition 4.18, it must approach $+\infty$. If $v(f) > v(x)$ then by Fact 4.20, $\lim_{x\to\infty} \Delta(f) = 0$. Since $\Delta(f) := f(x+1) - f(x)$ is definable in $\mathcal{M}$, it follows from Proposition 4.18 that it must be eventually 0 and therefore f is eventually affine.

If $v(f) < v(x)$ then by Fact 4.20, $v(f^{-1}) > v(x)$, where f^{-1} is taken to be the eventual compositional inverse of f, which is also definable in $\mathcal{M}$. Thus, as above, f^{-1} is eventually affine so also f is.

We are left with the case $v(f) = v(x)$. By Fact 4.20(3), the $\mathcal{M}$-definable function $\Delta(f)$ approaches a limit c in R. By Proposition 4.18, we have $\Delta(f)$ eventually constant, and thus, by o-minimality, f is eventually affine.

Thus, we showed so far that there exists a definable endomorphism $\lambda \in \Lambda_{\mathcal{M}}$ such that $f(x) = \lambda(x) + d$ for all $x > 0$ large enough. By Proposition 4.16, λ is $\varnothing$-definable. The set

$$\{x \in R : f(x) = \lambda(x) + d\}$$

is $\mathcal{M}$-definable and contains an unbounded ray so must be cobounded. □

Before the next proposition, we introduce a new notion.

Definition 4.22. Given $X \subseteq R^n$, let

$$\mathrm{Stab}_{\mathrm{bd}}(X) := \{a \in R^n : (a + X) \triangle X \text{ is bounded}\},$$

where $A \triangle B = A \cup B \setminus A \cap B$.

For a function f, *we let* $\Gamma(f)$ *denote its graph.*

By Proposition 4.3, if X is definable in $\mathcal{M}$ over A then so is $\mathrm{Stab}_{\mathrm{bd}}(X)$. The following facts are easy to verify:

Fact 4.23. (1) *For every* $X \subseteq R^n$, $\mathrm{Stab}_{\mathrm{bd}}(X)$ *is a subgroup of* $\langle R^n, +\rangle$.

(2) *If* $X \subseteq R^2$ *is the graph of an affine function* $f(x) = \lambda(x) + b$, *on a cobounded subset of* R, *then*

$$\mathrm{Stab}_{\mathrm{bd}}(X) = \Gamma(\lambda).$$

(3) *If a definable set* $X \subseteq R^2$ *is a finite union of graphs of affine functions, all of the form* $\lambda + d$ *for a fixed* λ, *and at least one of the functions is defined on an unbounded set, then* $\mathrm{Stab}_{\mathrm{bd}}(X) = \Gamma(\lambda)$.

The following statement would have been immediately true if definable sets in $\mathcal{M}$ admitted definable cell decomposition (with respect to the ambient ordering).

Proposition 4.24. *Assume that* $X \subseteq R^2$ *is* $\mathcal{M}$*-definable over* A, *and* $\dim(X) \leq 1$. *Assume that there exists an* $\mathcal{R}_{\mathrm{omin}}$*-definable endomorphism* $\lambda : R \to R$, *and some* $a, d \in R$ *such that graph of* $\lambda(x) + d \restriction (a, \infty)$ *is contained in* X. *Then* λ *is* $\mathcal{M}$*-definable* (*necessarily over* $\varnothing$).

Proof. Recall that $\mathscr{A}(X)$, the affine part of X is $\mathcal{M}$-definable over A. For large enough a, it contains $\Gamma(\lambda + d \restriction (a, \infty))$. So, without loss of generality, $X = \mathscr{A}(X)$.

We define for each $x, y \in X$, the relation $x \sim y$ if there exist open sets $U, V \ni 0$ in R^2 such that

$$(y - x) + (x + U \cap X) = y + V \cap X.$$

Said differently, up to translation, X has the same germ at x and at y. Because a basis for the R^2 topology is definable in $\mathcal{M}$, the relation $\sim$ is definable in $\mathcal{M}$.

Notice that for x large enough, all elements on $\Gamma(\lambda + d) \cap X$ are in the same $\sim$-class, so we may replace X by this $\sim$-class, which is $\mathcal{M}$-definable.

Thus, we may assume that all elements of X are $\sim$-equivalent, and X contains $\Gamma(\lambda + d \restriction (a, \infty))$. It follows that X is contained in finitely many translates of the graph of λ. Applying Fact 4.23(3), we conclude that $\mathrm{Stab}_{\mathrm{bd}}(X)$ is exactly the graph of λ, and thus the function $\lambda(x)$ is $\mathcal{M}$-definable. By Proposition 4.16, λ is $\varnothing$-definable. $\square$

Definable subsets of $\boldsymbol{R^2}$**.** The next result is the main structure theorem of the paper.

Theorem 4.25. *Under our standing assumptions on* $\mathcal{M}$, *assume that* $X \subseteq R^2$ *is definable in* $\mathcal{M}$ *over a parameter set* $A \subseteq R$, *with* $\dim(X) \leq 1$. *Then there are* $\lambda_1, \ldots, \lambda_r \in \Lambda_{\mathcal{M}}$ *and* $\mathcal{M}$*-definable finite sets* $D_i \subseteq R$, $i = 1, \ldots, r$, *and* $D \subseteq R$ *all defined over* A, *such that*

(i) *For every* $i = 1, \ldots, r$, *and* $d \in D_i$, $\Gamma(\lambda_i + d) \setminus X$ *is bounded* (*i.e., X contains the restriction of* $\lambda_i + d$ *to a cobounded set*).

(ii) *For every* $d \in D$, $(\{d\} \times R) \setminus X$ *is bounded.*

(iii) *The set*

$$X \setminus \left(\bigcup_{i=1}^{r} \bigcup_{d \in D_i} \Gamma(\lambda_i + d) \cup \bigcup_{d \in D} \{d\} \times R \right)$$

is bounded in R^2.

Proof. If X is bounded then there is nothing to prove, so we assume $\dim(X) = 1$ and X is unbounded. By the cell decomposition theorem in o-minimal structures, X can be decomposed into a finite union of cells of dimension 0 and 1. However, these cells are not in general definable in $\mathcal{M}$.

Assume first that X contains the graph of a function $f : (a, +\infty) \to R$, and let $\Psi(x, y)$ be the $\mathcal{M}$-formula that defines X.

Case (i): f is bounded at ∞.

In this case we prove a general statement:

Claim 4.26. *If* $\dim X \leq 1$ *and* X *contains the graph of a bounded function* $f : (a, \infty) \to R$ *then* f *is eventually constant.*

Proof. By o-minimality, $\lim_{x\to+\infty} f(x) = L$ for some $L \in R$.

By our standing assumption, $<\restriction(0, a_0)$ is $\mathcal{M}$-definable, for some $a_0 > 0$, and thus $<$ is definable on every interval of length $\leq a_0$. Let $X_L := R \times [L-a_0, L+a_0]) \cap X$. By o-minimality, there exists $m \in \mathbb{N}$ such that for all large enough $a \in R$, we have $|X_a| \leq m$. The set $Z = \{a \in R : |X_a| \leq m\}$ is definable in $\mathcal{M}$ and unbounded, so we may replace X_L by $X_L \cap Z \times R$, containing the graph of f. We call it X_L again.

Using the restricted order, we can partition X_L, definably in $\mathcal{M}$, into finitely many graphs of functions $g_1, g_2, \ldots, g_k$, $k \leq m$. For instance, we let

$$g_1(x) = \min\{y \in [L - a_0, L + a_0] : \langle x, y\rangle \in X_L\}$$

and continue similarly to obtain the other g_i's. For x large enough, the function f is one of those g_i's, and therefore it is $\mathcal{M}$-definable. Using Proposition 4.18 we get that f is eventually constant. □

Case (ii): $\lim_{x\to+\infty} f(x) = +\infty$.

We recall the proof of Lemma 4.21, and consider three cases: $v(f) > v(x)$, $v(f) < v(x)$ and $v(f) = v(x)$ (remembering though that we do not know yet that f is an $\mathcal{M}$-definable function).

Assume first that $v(f) > v(x)$. By Fact 4.20, $f(x + 1) - f(x) \to 0$, as $x \to \infty$. We want to capture $\Delta(f) = f(x + 1) - f(x)$ within an $\mathcal{M}$-definable set.

The formula

$$\varphi(x, y) := \exists z_1 \exists z_2\, (\Psi(x + 1, z_1) \wedge \Psi(x, z_2) \wedge (y = z_1 - z_2))$$

defines in $\mathcal{M}$ a new subset of R^2 — call it $\Delta(X)$ — which contains the graph of $\Delta(f)$ (but possibly more functions).

We first note that $\dim(\Delta(X)) = 1$. Indeed, for $a \in R$, $\Delta(X)_a$ is infinite if either X_a or X_{a+1} is infinite. Since only finitely many X_a's are infinite the same is true for $\Delta(X)$. Thus, the graph of $\Delta(f)$ is contained in the one-dimensional $\mathcal{M}$-definable set $\Delta(X)$, so by Claim 4.26, $\Delta(f)$ must be eventually constant, implying that f is eventually affine.

Assume now that $v(f) < v(x)$. The formula $\Upsilon(x, y) := \Psi(y, x)$ defines in $\mathcal{M}$ a new set X^{-1} containing the graph of f^{-1} (a partial function). The graph of f^{-1} is still contained in X^{-1} and we have $v(f^{-1}) > v(x)$. Thus, applying the case we already handled, we see that f^{-1}, and hence also f is eventually affine.

We are left with the case $v(f) = v(x)$. Using Fact 4.20(3), the function $\Delta(f)$ tends to a constant. Thus, as above, we may use the $\mathcal{M}$-definable set $\Delta(X)$ to deduce that $\Delta(f)$ is eventually constant and thus f is eventually affine.

So far we handled all cases where the unbounded cell in X is the graph of some function on a ray (a, ∞). The same reasoning applies to rays $(-\infty, a)$. Applying this reasoning to X^{-1}, we obtain in addition those functions which are eventually constant in X^{-1}, namely sets of the form $\{d\} \times R$ whose intersection with X is co-unbounded in $\{d\} \times R$. The set of all such d is clearly definable over A.

To summarize, we showed that every unbounded cell in X is either contained in the graph of an eventually affine function f definable in $\mathcal{M}$, or in $\{d\} \times R$ for some d. By Proposition 4.24, the function f has the form $\lambda(x) + d$ for $\lambda \in \Lambda_{\mathcal{M}}$. Thus, we have $\lambda_1, \dots, \lambda_k \in \Lambda_{\mathcal{M}}$, and for each such $i = 1, \dots, k$, the set D_i of $d \in R$ such that $\Gamma(\lambda_i + d) \cap X$ is unbounded, is $\mathcal{M}$-definable over A, and must be finite. For every such d, $\Gamma(\lambda_i + d) \setminus X$ is bounded.

The above proof handles all unbounded cells, so the set

$$X \setminus \left(\bigcup_{i=1}^{r} \bigcup_{d \in D_i} \Gamma(\lambda_i + d) \cup \bigcup_{d \in D} \{d\} \times R \right)$$

is bounded. □

The algebraic closure and definable closure in strongly bounded structures. Even though the full ordering on R is not definable, we can still prove:

Theorem 4.27. *The algebraic closure in $\mathcal{M}$ equals the definable closure. Moreover, if $a \in \operatorname{acl}_{\mathcal{M}}(\bar{b})$ then a is in the $\mathcal{L}_{\mathrm{bd}}$-definable closure of $\bar{b}$.*

Proof. We use acl, dcl and $\operatorname{acl}_{\mathrm{bd}}$, $\operatorname{dcl}_{\mathrm{bd}}$ to denote the corresponding operations in $\mathcal{M}$ and $\mathcal{M}_{\mathrm{bd}}$, respectively. We prove by induction on n that if $a \in \operatorname{acl}(b_1, \dots, b_n)$, for some $a, b_i \in R$, then $a \in \operatorname{dcl}_{\mathrm{bd}}(b_1, \dots b_n)$.

We first handle the case $n = 0$, namely $a \in \operatorname{acl}(\varnothing)$. In this case, there is a finite $\varnothing$-definable set $A \subseteq M$ such that $a \in A$. Viewing the set A in $\mathcal{R}_{\mathrm{omin}}$, we can order the elements $a_1 < \cdots < a_n$. The interval (a_1, a_n) is a $\varnothing$-interval, and $< \restriction (a_1, a_n)$ is $\mathcal{M}_{\mathrm{bd}}$-definable over $\varnothing$, so each a_i is in $\operatorname{dcl}_{\mathrm{bd}}(\varnothing)$.

We proceed by induction, and assume that we proved the result for $n-1$. Assume now that $a \in \operatorname{acl}(b_1, \dots, b_{n-1}, b_n)$. Let $X \subseteq R^{n+1}$ be a $\varnothing$-definable set such that

$\langle b_1, \dots, b_n, a\rangle$ and $X_{b_1,\dots,b_n}$ has size m. Without loss of generality, for every b'_n, the set $X_{b_1,\dots,b_{n-1},b'_n}$ has size m.

Let $b' = (b_1, \dots, b_{n-1})$ and consider the set $X_{b'} = \{\langle x, y\rangle \in R^2 : \langle b', x, y\rangle \in X\}$. By our assumption, $\dim(X_{b'}) \leq 1$ and $\langle b_n, a\rangle \in X_{b'}$.

We now apply Theorem 4.25. We obtain finitely many $\varnothing$-definable endomorphisms $\lambda_1, \dots, \lambda_k \in \Lambda_{\mathcal{M}}$ and for each $i = 1, \dots, k$, we have a b'-definable finite set A_i such that

$$X^{\mathrm{bd}}_{b'} = X_{b'} \setminus \left(\bigcup_{i=1}^{k} \bigcup_{d\in A_i} \Gamma(\lambda_i + d)\right)$$

is bounded in R^2.

Since $|b'| = n-1$, it follows by induction that every $d \in A_i$ is in $\mathrm{dcl}_{\mathrm{bd}}(b')$. Assume first that $\langle b_n, a\rangle$ is in the graph of one of the $\lambda_i + d$, $d \in A_i$, namely $a = \lambda_i(b_n) + d$. Because λ_i is $\varnothing$-definable and $d \in \mathrm{dcl}_{\mathrm{bd}}(b')$ it follows that $a \in \mathrm{dcl}_{\mathrm{bd}}(b_1, \dots, b_n)$.

We are left with the case $\langle b_n, a\rangle \in X^{\mathrm{bd}}_{b'}$. The set $X^{\mathrm{bd}}_{b'}$ is b'-definable so we may assume that $X_{b'} = X^{\mathrm{bd}}_{b'}$ is bounded (but possibly not $\varnothing$-bounded). Let π_1, π_2 be the projection of $X_{b'}$ onto the first and second coordinates. Each of these is a finite union of points and pairwise disjoint bounded open intervals. Let

$$\pi_1(X_{b'}) = F_1 \cup \bigcup_{i=1}^{k}(a_i, b_i) \quad \text{for } F_1 \text{ finite and } a_1 < b_1 < \cdots < a_k < b_k,$$

and

$$\pi_2(X_{b'}) = F_2 \cup \bigcup_{j=1}^{r}(c_j, d_j) \quad \text{for } F_2 \text{ finite and } c_1 < d_1 < \cdots < c_r < d_r.$$

By Theorem 4.9, there is a fixed $K \in \mathrm{dcl}(\varnothing)$ such that for all $i, = 1, \dots, k$ and $j = 1, \dots, r$, we have $b_i - a_i, d_j - c_j \leq K$.

By Lemma 4.8, the sets $\{a_i\}, \{b_i\}, \{c_j\}, \{d_j\}$ are all finite and $\mathcal{M}$-definable over b', and thus, by induction each of these endpoints is in $\mathrm{dcl}_{\mathrm{bd}}(b')$. Assume that $\langle b_n, a\rangle \in X \cap (a_i, b_i) \times (c_j, d_j)$ for some $i = 1, \dots, k$ and $j = 1, \dots, r$. We replace X by the b'-definable set $X_1 = X - \langle a_i, c_j\rangle \cap (0, b_i - a_i) \times (0, d_j - c_j) \subseteq (0, K)^2$. Notice that $\langle b_n - a_i, a - c_j\rangle \in X'$, and the fiber in X' over $b_n - a_i$ is finite. Because the ordering on $(0, K)$ is $\mathcal{M}_{\mathrm{bd}}$-definable over $\varnothing$, we have $a - c_j \in \mathrm{dcl}_{\mathrm{bd}}(b', b_n - a_i)$, but since $a_i, c_j \in \mathrm{dcl}_{\mathrm{bd}}(b')$ we have $a \in \mathrm{dcl}_{\mathrm{bd}}(b', b_n)$. This ends the proof that $\mathrm{acl} = \mathrm{dcl}_{\mathrm{bd}}$ in $\mathcal{M}$. □

Definable subsets of R^n. We are now ready to prove the main theorem, under the assumptions outlined on p. 387.

Theorem 4.28. *If $X \subseteq R^n$ is $\mathcal{M}$-definable over $A \subseteq R$ then X is definable in $\mathcal{M}_{\mathrm{bd}}$ over A.*

Proof. It is sufficient to prove the result in $\mathcal{N} \succ \mathcal{M}$, so by replacing $\mathcal{R}_{\text{omin}}$ (thus also its reducts) by a sufficiently saturated extension, we may assume that $\mathcal{M}$ is ω-saturated.

We prove the result by induction on n. For $X \subseteq R$, the set X is either bounded or cobounded, so we may assume that it is bounded. Thus, it can be written as a disjoint union

$$(a_1, b_1) \cup \cdots \cup (a_n, b_n) \cup F,$$

with $a_1 < b_1 < \cdots < a_n < b_n$ and F finite. By Lemma 4.8, each a_i and b_i is in $\text{acl}_{\mathcal{M}}(A)$, so by Theorem 4.27, it belong to $\text{dcl}_{\text{bd}}(A)$. Similarly, $F \subseteq \text{dcl}_{\text{bd}}(A)$. By Theorem 4.9, there is $K \in \text{dcl}_{\text{bd}}(\varnothing)$ such that all intervals (a_i, b_i) are of length at most K. But then each interval $(0, b_i - a_i)$ is contained in a $\varnothing$-interval, hence definable in $\mathcal{M}_{\text{bd}}$ over A, so also (a_i, b_i) is $\mathcal{M}_{\text{bd}}$-definable over A. It follows that X is definable in $\mathcal{M}_{\text{bd}}$.

We now use induction on n. Given $X \subseteq R^{n+1}$ that is $\mathcal{M}$-definable over A, we consider, for each $t \in R^n$, the set

$$X_t = \{b \in R : \langle t, b\rangle \in X\} \subseteq R.$$

By the case $n = 1$, each X_t is $\mathcal{M}_{\text{bd}}$-definable over At. Thus, by compactness and saturation, we can find $\mathcal{L}_{\text{bd}}$-formulas over A, $\varphi_1(t, x), \ldots, \varphi_k(t, x)$ such that for every $t \in R^n$, one of the $\varphi_i(t, x)$ defines X_t. Let

$$T_i = \big\{t \in R^n : \exists x \,\big(\langle t, x\rangle \in X \wedge \forall x\, (x \in X_t \leftrightarrow \varphi_i(t, x))\big)\big\}.$$

The set T_i is $\mathcal{M}$-definable, over A, and thus, by induction, it is $\mathcal{M}_{\text{bd}}$-definable over A by some $\psi_i(t)$. The formula $\varphi_i(t, x) \wedge \psi_i(t)$ defines $X \cap T_i \times R$, so X is definable in $\mathcal{M}_{\text{bd}}$ over A. □

A comment on failure of definable choice in strongly bounded $\mathcal{M}$. Recall that a structure $\mathcal{M}$ has definable choice if for every definable family $\{X_t : t \in T\}$ of sets, there is a definable function $f : T \to \bigcup X_t$ such that $f(t) \in X_t$ and if $t_1 = t_2$ then $f(t_1) = f(t_2)$. Equivalently, every definable equivalence relation has a definable set of representatives. This fails in strongly bounded $\mathcal{M}$, because the relation $xEy \Leftrightarrow y = -x$ on R cannot have a definable set of representatives. If it did then it would contain either a positive or a negative ray (without its inverse).

We believe that elimination of imaginaries similarly fails.

5. Conclusion: The proof of Theorem 1.2

We are now ready to collect the results proved thus far in order to prove Theorem 1.2.

Recall that we want to prove that the only reducts between $\mathcal{R}_{\text{lin}}$ and $\mathcal{R}_{\text{alg}}$ are as follows:

$$\mathcal{R}_{\text{alg}} = \langle R; +, \cdot, < \rangle,$$
$$\mathcal{R}_{\text{sb}} = \langle R; +, <, \Lambda_R, \mathfrak{B} \rangle,$$
$$\mathcal{R}_{\text{semi}} = \langle R; +, <, \Lambda_R \rangle, \qquad \mathcal{R}_{\text{bd}} = \langle R; +, <^*, \Lambda_R, \mathfrak{B} \rangle,$$
$$\mathcal{R}^*_{\text{lin}} = \langle R; +, <^*, \Lambda_R \rangle,$$
$$\mathcal{R}_{\text{lin}} = \langle R; +, \Lambda_R \rangle.$$

First, we note that using [Edmundo 2000] we can generalize [Peterzil 1993, Theorem 1.1] from $\mathbb{R}$ to arbitrary real closed fields, and show:

Fact 5.1. *Let R be a real closed field. The only reduct between $\mathcal{R}_{\text{semi}}$ and $\mathcal{R}_{\text{alg}}$ is $\mathcal{R}_{\text{sb}}$.*

Proof. Assume that $\mathcal{M}$ is a reduct of R_{alg} which properly expands $\mathcal{R}_{\text{semi}}$. By [Edmundo 2000, Fact 1.6], either $\mathcal{M}$ is a reduct of $\mathcal{R}_{\text{sb}}$ or a real closed field $F = \langle R; \oplus, \odot \rangle$ whose universe R is definable in $\mathcal{M}$. Assume the latter, and then since the field is semialgebraic then, again by [Peterzil 1993, Corollary 2.4], every semialgebraic subset of R is definable in F and hence in $\mathcal{M}$. Thus, $\mathcal{M} \doteq \mathcal{R}_{\text{alg}}$.

If $\mathcal{M}$ is a reduct of $\mathcal{R}_{\text{sb}}$ which is not semilinear then by Theorem 3.2, every bounded R-semialgebraic set is definable in $\mathcal{M}$, and thus $\mathcal{M} \doteq \mathcal{R}_{\text{sb}}$. ☐

We now consider an arbitrary reduct $\mathcal{M}$ of $\mathcal{R}_{\text{alg}}$. Our goal is to show that $\mathcal{M}$ is one of the reducts in the above list.

First, if $\mathcal{M}$ is stable then by Claim 2.2, $\mathcal{R}_{\text{lin}} \doteq \mathcal{M}$. If $\mathcal{M}$ is unstable then by Theorem 2.1, $<^*$ is definable in $\mathcal{M}$. So $\mathcal{R}^*_{\text{lin}} \dot\subseteq \mathcal{M}$. So, we may assume that $<^*$ is definable in $\mathcal{M}$, i.e., $\mathcal{R}^*_{\text{lin}} \dot\subseteq \mathcal{M}$.

Case 1: $\mathcal{M}$ is strongly bounded and $\mathcal{M} \dot\subseteq \mathcal{R}_{\text{semi}}$.

We claim that $\mathcal{M} \doteq \mathcal{R}^*_{\text{lin}}$. Indeed, because $\mathcal{M}$ is strongly bounded then, by Theorem 4.5, $\mathcal{M} \doteq \mathcal{M}_{\text{bd}}$. Because $\mathcal{M} \dot\subseteq \mathcal{R}_{\text{semi}}$, every $\mathcal{M}$-definable set is semilinear, and in particular this is true for each of the $\varnothing$-bounded sets in $\mathcal{M}_{\text{bd}}$. However, it is easy to verify that every bounded semilinear set is definable in $\mathcal{R}^*_{\text{lin}}$, so the whole structure $\mathcal{M}_{\text{bd}}$ is a reduct of $\mathcal{R}^*_{\text{lin}}$, and thus so is $\mathcal{M}$ as well. The converse $\mathcal{R}^*_{\text{lin}} \dot\subseteq \mathcal{M}$ is already assumed.

Case 2: $\mathcal{M}$ is strongly bounded and $\mathcal{M} \dot\nsubseteq \mathcal{R}_{\text{semi}}$.

We claim that $\mathcal{M} \doteq \mathcal{R}_{\text{bd}}$. As in Case 1, every $\mathcal{M}$-definable set is definable in $\mathcal{M}_{\text{bd}}$. Because $\mathcal{M}$ is a reduct of $\mathcal{R}_{\text{alg}}$ then $\mathcal{M}_{\text{bd}}$ is a reduct of $\mathcal{R}_{\text{bd}}$ and so $\mathcal{M} \dot\subseteq \mathcal{R}_{\text{bd}}$. By the assumption that $\mathcal{M} \dot\nsubseteq \mathcal{R}_{\text{semi}}$, we know that there is an $\mathcal{M}$-definable semialgebraic set which is not semilinear, so by Theorem 3.2, we get that every bounded semialgebraic set is definable in $\mathcal{M}$, hence $\mathcal{R}_{\text{bd}} \dot\subseteq \mathcal{M}$.

Next we assume that $\mathcal{M}$ is *not* strongly bounded.

Case 3: $\mathcal{M}$ is not strongly bounded and $\mathcal{M} \dot{\subseteq} \mathcal{R}_{\text{semi}}$.

By Lemma 2.3, the linear order $<$ is definable in $\mathcal{M}$, so, since $\mathcal{R}^*_{\text{semi}} \dot{\subseteq} \mathcal{M}$, we have $\mathcal{R}_{\text{semi}} \doteq \mathcal{M}$.

Case 4: $\mathcal{M}$ is not strongly bounded and $\mathcal{M} \dot{\nsubseteq} \mathcal{R}_{\text{semi}}$.

As in Case 3, the linear order $<$ is definable in $\mathcal{M}$, so $\mathcal{R}_{\text{semi}} \dot{\subseteq} \mathcal{M}$. So we know that $\mathcal{M}$ is a reduct of $\mathcal{R}_{\text{alg}}$ which properly expands $\mathcal{R}_{\text{semi}}$. By Fact 5.1, either $\mathcal{M} \doteq \mathcal{R}_{\text{alg}}$ or $\mathcal{M} \doteq \mathcal{R}_{\text{bd}}$.

This completes the proof that if $\mathcal{M}$ is a reduct of $\mathcal{R}_{\text{alg}}$ expanding $\mathcal{R}_{\text{lin}}$, then it is one of the reducts in the above diagram.

It is left to see that all reducts in the above diagram are distinct. Because $\mathcal{R}_{\text{lin}}$ is stable and $\mathcal{R}^*_{\text{lin}}$ is unstable, these two are distinct. Also, the fact that $\mathcal{R}^*_{\text{lin}}$ and $\mathcal{R}_{\text{bd}}$ are distinct is easy to verify (e.g., the unit circle is definable in $\mathcal{R}_{\text{bd}}$ but not in $\mathcal{R}^*_{\text{lin}}$). The fact that $\mathcal{R}_{\text{bd}}$ is different than $\mathcal{R}_{\text{sb}}$ and $\mathcal{R}_{\text{semi}}$ follows from the next lemma.

Lemma 5.2. *Let R be a real closed field. If $\mathcal{B}^*$ is any collection of bounded subsets of R^n, $n \in \mathbb{N}$, then $<$ is not definable in $\mathcal{M} = \langle R; +, \Lambda_R, \mathcal{B}^* \rangle$.*

Proof. We use a similar idea to [Peterzil 1992] Assume towards a contradiction that $<$ is definable in $\mathcal{M}$, and let $\mathcal{N} = \langle R : +, <, \Lambda_R, \mathfrak{B}^* \rangle$.

Let $\psi(x, y, \bar{a})$, $\bar{a} \in R$ be the $\mathcal{M}$-formula that defines $<$. Namely,

$$\mathcal{N} \models \forall x \forall y \, (\psi(x, y, \bar{a}) \leftrightarrow x < y).$$

Let $\widetilde{\mathcal{N}} = \langle \widetilde{R}; +, <, \Lambda_R, \mathfrak{B}^* \rangle \succ \mathcal{N}$ be an $|N|^+$-saturated elementary extension whose reduct to the $\mathcal{M}$-language is $\widetilde{\mathcal{M}}$. It follows that $\psi(x, y, \bar{a})$ defines $<$ in $\widetilde{\mathcal{N}}$ as well.

We show that there is an automorphism of $\widetilde{\mathcal{M}}$ which fixes $\bar{a}$, thus leaving $\psi(\widetilde{R} \times \widetilde{R}, \bar{a})$ invariant, and yet not respecting $<$, leading to a contradiction.

The group $\langle \widetilde{R}, + \rangle$ is a vector space over R. We define an R-vector subspace of $\widetilde{R}$ by

$$\mathcal{A} = \{x \in \widetilde{R} : \exists \alpha \in R \, (|x| < \lambda_\alpha(1))\}.$$

So, by Zorn's lemma, there exists an R-vector space $V \subseteq \widetilde{R}$ such that $\widetilde{R} = \mathcal{A} \oplus V$, and by the saturation assumption, V is nontrivial. Now we define the following automorphism of the R-vector space $\widetilde{R}$: on $\mathcal{A}$ we define $\tau_1(v) = v$, on $\langle V, + \rangle$ we define $\tau_2(v) = -v$, and we let $\tau : \widetilde{R} \to \widetilde{R}$ be

$$\tau(v_1 + v_2) = \tau_1(v_1) + \tau_2(v_2) = v_1 - v_2.$$

This automorphism fixes all elements in $\mathcal{A}$ and in particular fixes all sets in $\mathfrak{B}^*$ pointwise, but does not respect $<$ (as positive elements in V are sent to negative ones). In model theoretic language τ is an automorphism of the structure $\widetilde{\mathcal{M}}$ which fixes $\bar{a}$ (since $\bar{a} \in \mathcal{A}$). However, τ does not preserve $<$, contradiction. ☐

This completes the proof of Theorem 1.2. ☐

Appendix: The proof of Fact 3.1

Fact 3.1. *Let R be a real closed field and $X \subseteq R^n$ a definable set in an o-minimal expansion of $\langle R; <, +, \cdot \rangle$. If X is not definable in $\mathcal{R}_{\text{semi}}$ then, in the structure $\mathcal{M} = \langle R; <^*, +, \Lambda_R, X \rangle$ there exists a definable bounded set which is not definable in $\mathcal{R}_{\text{semi}}$.*

Proof. We believe that this is known so we shall be brief. We prove the result by induction on $\dim(X)$, where the case $\dim X = 0$ is trivially true. Consider the affine part of X, $\mathscr{A}(X)$, which is definable in $\mathcal{M}$.

Assume first $\mathscr{A}(X)$ is not dense in X. Then there is an open box $U \subseteq R^n$ such that $U \cap X \neq \varnothing$ and $U \cap \mathscr{A}(X) = \varnothing$. We claim that $U \cap X$ is not semilinear. Indeed, if it were then $\mathscr{A}(U \cap X)$ must be nonempty, but because $U \cap X$ is relatively open in X then

$$\mathscr{A}(U \cap X) = U \cap \mathscr{A}(X) = \varnothing,$$

a contradiction.

Thus, $U \cap X$ above is not semilinear. and this gives the desired box when $\mathscr{A}(X)$ is not dense in X.

We assume then that $\mathscr{A}(X)$ is dense in X, and consider two cases: $\mathscr{A}(X)$ is either semilinear or not. If it were semilinear then necessarily $X \setminus \mathscr{A}(X)$ is not semilinear, and because of the density assumption, $\dim(X \setminus \mathscr{A}(X)) < \dim(X)$ and we can finish by induction.

Thus, we are left with the case that $\mathscr{A}(X)$ is not semilinear. For simplicity, we may assume now that $X = \mathscr{A}(X)$. We recall the $\mathcal{M}$-definable relation $a \sim b$ from the proof of Proposition 4.24, defined by letting $a \sim b$ if X has the same germ at a and b, up to translation.

Because $X = \mathscr{A}(X)$, each $\sim$-class is open in X, and thus there are finitely many classes, at least one of which is not semilinear. Thus, we may assume that $X = \mathscr{A}(X)$ consists of a single $\sim$-class. It follows that there is some R-subspace $L \subseteq R^n$, $\dim L = \dim X$, such that X is contained in a finite union of cosets of L. Thus each definably connected component of X is contained in a single such coset of L.

Each L is definable in $\mathcal{M}$ using Λ_R, so the intersection of X with each of these cosets is definable in $\mathcal{M}$. One of these intersections is not semilinear, so we may assume that $X \subseteq c + L$ for some c. Because $\dim X = \dim L$, and $\mathscr{A}(X) = X$, then X is open in $c + L$. We claim that $\mathrm{Fr}(X) \subseteq c + L$ is not semilinear: Indeed, $\mathrm{Fr}(X)$ is a closed subset of $c + L$, and X consists of finitely many components of $c + L \setminus \mathrm{Fr}(X)$. If $\mathrm{Fr}(X)$ were semilinear then each of its components would also be, so X would be semilinear.

Thus, $\mathrm{Fr}(X)$ is not semilinear, and definable in $\mathcal{M}$. By o-minimality,

$$\dim(\mathrm{Fr}(X)) < \dim(X).$$

Therefore, by induction we may find an $\mathcal{M}$-definable bounded set which is not semilinear. $\square$

In fact, a stronger result is true: If $X \subseteq R^n$ is definable in an o-minimal expansion of the field R and not semilinear, then there is some bounded open box $U \subseteq R^n$ such that $U \cap X$ is not semilinear (we omit the proof here as we do not need it). Notice that this last statement fails if we replace "not semilinear" by "not semialgebraic", as Rolin's example from [Le Gal and Rolin 2009] shows: There exists a definable function $f : \mathbb{R} \to \mathbb{R}$ in an o-minimal expansion of the real field such that the restriction of f to every bounded interval is semialgebraic but f itself is not semialgebraic.

Acknowledgments

This article is based on results from the M.Sc. thesis of the first author, as part of her studies at the University of Haifa. We thank Itay Kaplan and Assaf Hasson for reading an earlier version of this work and commenting on it.

References

[Belegradek 2004] O. Belegradek, "Semi-bounded relations in ordered modules", *J. Symbolic Logic* **69**:2 (2004), 499–514. MR Zbl

[Conant 2018] G. Conant, "There are no intermediate structures between the group of integers and Presburger arithmetic", *J. Symb. Log.* **83**:1 (2018), 187–207. MR Zbl

[Edmundo 2000] M. J. Edmundo, "Structure theorems for o-minimal expansions of groups", *Ann. Pure Appl. Logic* **102**:1-2 (2000), 159–181. MR Zbl

[Hasson et al. 2010] A. Hasson, A. Onshuus, and Y. Peterzil, "Definable one dimensional structures in o-minimal theories", *Israel J. Math.* **179** (2010), 297–361. MR Zbl

[Hrushovski and Pillay 1987] U. 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

[Le Gal and Rolin 2009] O. Le Gal and J.-P. Rolin, "An o-minimal structure which does not admit C^∞ cellular decomposition", *Ann. Inst. Fourier* (*Grenoble*) **59**:2 (2009), 543–562. MR Zbl

[Loveys and Peterzil 1993] J. Loveys and Y. Peterzil, "Linear o-minimal structures", *Israel J. Math.* **81**:1-2 (1993), 1–30. MR Zbl

[Marker et al. 1992] D. Marker, Y. Peterzil, and A. Pillay, "Additive reducts of real closed fields", *J. Symbolic Logic* **57**:1 (1992), 109–117. MR Zbl

[Miller and Starchenko 1998] C. Miller and S. Starchenko, "A growth dichotomy for o-minimal expansions of ordered groups", *Trans. Amer. Math. Soc.* **350**:9 (1998), 3505–3521. MR Zbl

[Peterzil 1992] Y. Peterzil, "A structure theorem for semibounded sets in the reals", *J. Symbolic Logic* **57**:3 (1992), 779–794. MR Zbl

[Peterzil 1993] Y. Peterzil, "Reducts of some structures over the reals", *J. Symbolic Logic* **58**:3 (1993), 955–966. MR Zbl

[Peterzil 2009] Y. Peterzil, "Returning to semi-bounded sets", *J. Symbolic Logic* **74**:2 (2009), 597–617. MR Zbl

[Peterzil and Starchenko 1998] Y. Peterzil and S. Starchenko, "A trichotomy theorem for o-minimal structures", *Proc. London Math. Soc.* (3) **77**:3 (1998), 481–523. MR Zbl

[Pillay et al. 1989] A. Pillay, P. Scowcroft, and C. Steinhorn, "Between groups and rings", *Rocky Mountain J. Math.* **19**:3 (1989), 871–885. MR Zbl

Received 9 May 2022. Revised 19 Nov 2022.

HIND ABU SALEH:

hind.abu.94@gmail.com
Department of Mathematics, University of Haifa, Haifa, Israel

YA'ACOV PETERZIL:

kobi@math.haifa.ac.il
Department of Mathematics, University of Haifa, Haifa, Israel

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.405

Remarks around the nonexistence of difference closure

Zoé Chatzidakis

This paper shows that in general, difference fields do not have a difference closure. However, we introduce a stronger notion of closure (κ-closure), and show that every algebraically closed difference field K of characteristic 0, with fixed field satisfying a certain natural condition, has a κ-closure, and this closure is unique up to isomorphism over K.

Introduction

In this paper, a difference field is a field K with a distinguished automorphism σ. A difference field L is *difference closed* if every finite system of difference equations with coefficients in L which has a solution in a difference field extending L, already has a solution in L.

If K is a difference field, then a *difference closure of* K is a difference closed field containing K, and which K-embeds into every difference closed field containing K.

The algebra of difference fields was developed by Ritt, in analogy with the algebra of differential fields. It is well-known that any differential field of characteristic 0 has a differential closure, and that this differential closure is unique up to isomorphism over the field. In 2016, Michael Singer asked whether this result generalises to the context of difference fields. One of the main results of this paper is that it does not, even after imposing some natural conditions on the difference field K. We will show by two examples (Examples 1.3 and 1.4) that even the existence of a difference closure can fail.

There are several natural strengthenings of the notions of difference closed and difference closure (originating from model theory but having a natural algebraic translation), and we will show that these notions do satisfy existence and uniqueness of closure, provided we work over an algebraically closed difference field of characteristic 0 whose fixed subfield is large enough.

The theory of difference closed difference fields has been extensively studied, and is commonly denoted by ACFA. The proof of our result uses in an essential

Partially supported by ANR–13-BS01-0006 (ValCoMo), ANR–17-CE40-0026 (AGRUME) and AAPG2019 (ANR-DFG Geomod).
MSC2020: 03C50, 03C60, 03C98, 12Hxx.
Keywords: difference fields, prime models.

way the characteristic 0 hypothesis, as it allows us to use techniques of stability theory. They provide examples of structures which are stable over a predicate; see [13; 14] for definitions. The main result of the paper is as follows:

Theorem 3.14. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$, and let K be an algebraically closed difference field of characteristic* 0 *such that $F := \mathrm{Fix}(\sigma)(K)$ is pseudofinite and is κ-saturated. Then there is a κ-prime model of* ACFA *over K. Furthermore, it is unique up to isomorphism over K.*

Here is an algebraic translation of this result for $\kappa \geq \aleph_1$: Call a difference field $\mathcal{U}$ *κ-closed* if every system of $< \kappa$ difference equations over $\mathcal{U}$ which has a solution in some difference field extending $\mathcal{U}$ has a solution in $\mathcal{U}$. The field $\mathcal{U}$ is a *κ-closure* of the difference field K if it is κ-closed, contains K, and K-embeds into every κ-closed difference field containing K. Then Theorem 3.14 states, for $\kappa \geq \aleph_1$:

> *Let K be an algebraically closed difference field of characteristic* 0, *whose fixed field F is pseudofinite and such that every system of $< \kappa$ polynomial equations over F which has a solution in a regular extension of F already has a solution in F. Then K has a κ-closure, and it is unique up to K-isomorphism.*

It is unlikely that this result can be generalised to the characteristic p context, and in fact, I conjecture that unless the difference field K of characteristic $p > 0$ is of cardinality $< \kappa$ or is already κ-closed, then it does not have a κ-closure.

The paper is organised as follows. In Section 1 we discuss the problem and reformulate it in model-theoretic terms, and describe the two examples. In Section 2, we state the preliminary results we will need from difference algebra and model theory. Section 3 contains the proof of Theorem 3.14.

1. Discussion of the problems and the examples

1.1. ***Notation and conventions.*** All difference fields will be inversive, i.e., the endomorphism σ will be onto. Let K be a difference field, contained in some large difference field $\mathcal{U}$. If a is a tuple in $\mathcal{U}$, we denote by $K(a)_\sigma$ the difference field generated by a over K, i.e., the subfield $K(\sigma^i(a))_{i\in\mathbb{Z}}$ of $\mathcal{U}$. The algebraic and separable closure of a field L are denoted by L^{alg} and L^s, respectively, and $G(K)$ denotes the absolute Galois group of K, i.e., $\mathcal{G}al(K^s/K)$. If $A \subset \mathcal{U}$, then $\mathrm{acl}(A)$ denotes the smallest algebraically closed difference field containing A; it coincides with the model-theoretic algebraic closure of A for the theory ACFA [5, Proposition 1.7]. We denote by $\mathcal{L}$ the language $\{+, -, \cdot, 0, 1, \sigma\}$.

1.2. ***Translation into model-theoretic terms.*** Let K be a difference field. Recall that any complete theory extending the theory ACFA of difference closed difference fields is supersimple, unstable, of SU-rank ω, and does not eliminate quantifiers, but it eliminates imaginaries. It is extensively studied in [5]. The reason ACFA does not eliminate quantifiers is that given an automorphism σ of a field K, there may be several nonisomorphic ways of extending σ to K^{alg}. So, the first obvious obstacle to the existence of a difference closure is that, and a natural condition to impose is to assume that K is algebraically closed. There is another natural condition one needs to impose: if L is difference closed, then its fixed field

$$\text{Fix}(\sigma)(L) = \{a \in L \mid \sigma(a) = a\}$$

is pseudofinite. Moreover, every pseudofinite field can occur as the fixed field of some difference closed field [1]. Thus if L is the difference closure of a difference field K, then $\text{Fix}(\sigma)(L)$ must be prime over $\text{Fix}(\sigma)(K)$ (for its theory in the language of rings). Duret showed in [8] that any completion of the theory of pseudofinite fields has the independence property. From his proof one extracts easily the fact that nonalgebraic types are nonisolated, and this forces us to require in case K is countable that $\text{Fix}(\sigma)(K)$ be pseudofinite in order to hope to have a difference closure. The case when K is uncountable is a little more complicated, the question is addressed and solved in [3].

It is therefore reasonable to make the following two assumptions:

K is algebraically closed, and $\text{Fix}(\sigma)(K)$ *is pseudofinite.*

But even this is not enough. To show this does not suffice, what we need to do is the following:

Exhibit a difference field K satisfying the above two conditions, and a finite system of difference equations which does not have a solution in K, and such that any finite strengthening of this system has several completions.

This looks easy, since even our stable types are only superstable, not ω-stable. However, the first obvious examples do not satisfy the first condition. Here is a more involved example, taken from [5, Example 6.7]:

Example 1.3 (an example in characteristic 0). Let k be a countable pseudofinite field of characteristic 0 containing $\mathbb{Q}^{\text{alg}}$, and consider $K = (k^{\text{alg}}, \sigma)$, where σ is a (topological) generator of $\mathcal{G}al(K/k)$. We consider the elliptic curve J_a, with j-invariant $a \notin K$, and which is defined by

$$y^2 + xy = x^3 - \frac{36}{a-1728}x - \frac{1}{a-1728}.$$

We let A' be a cyclic subgroup of J_a of order p^2, $A = [p]A'$ and a_1 the j-invariant of the elliptic curve J_a/A, a_2 the j-invariant of the elliptic curve J_a/A'. Then the map $\mathbb{Q}^{\text{alg}}(a, a_1) \to \mathbb{Q}^{\text{alg}}(a_1, a_2)$ which is the identity on $\mathbb{Q}^{\text{alg}}$ and sends (a, a_1) to (a_1, a_2) extends to a field automorphism of $\mathbb{Q}(a)^{\text{alg}}$, which in turns extends to an automorphism of $K(a)^{\mathrm{alg}}$ which agrees with σ on K. Let $\Phi(x, x_1, x_2)$ be the finite system of polynomial equations which describe the algebraic locus of (a, a_1, a_2) over K; in the notation of [11, Chapter 5, §3] (see in particular Theorem 5), $\Phi(x, x_1, x_2)$ can be written as

$$\Phi_{p^2}(x_2, x) = \Phi_p(x_1, x) = \Phi_p(x_2, x_1) = 0.$$

(The equation $\Phi_n(y, x) = 0$ says that y is the j-invariant of the quotient of the elliptic curve J_x with j-invariant x by a cyclic subgroup of order n.) We now consider the formula $\psi(x)$ given by $\Phi(x, \sigma(x), \sigma^2(x)) \wedge \sigma(x) \neq x$. Let b be any solution of $\psi(x)$. Note that necessarily, the kernel of the map $J_b \to J_{\sigma^n(b)}$ for $n > 0$ is cyclic of order p^n. Indeed, note that $\sigma^n(b)$ satisfies ψ for every n; hence, the kernel of the map $J_{\sigma^n(b)} \to J_{\sigma^{n+2}(b)}$ is cyclic of order p^2, and this map is the composite of the two maps $J_{\sigma^n(b)} \to J_{\sigma^{n+1}(b)}$ and $J_{\sigma^{n+1}(b)} \to J_{\sigma^{n+2}(b)}$, which both have kernel of order p. An easy induction then gives the result.

As $\sigma(b) \neq b$, we know that b is transcendental. Hence the curve J_b is not of CM-type, its endomorphism group is isomorphic to $\mathbb{Z}$, and therefore J_b is not isomorphic to any of its quotients by finite cyclic subgroups; see, e.g., [15, Section C.11]. Therefore, the elements $b, \sigma(b), \sigma^2(b), \ldots$ are all distinct, and $b \notin K$. Furthermore, the isomorphism type of $K(b)_\sigma$ over K is determined by $\Phi(b, \sigma(b), \sigma^2(b))$, because as we saw above, the kernel of the map $J_b \to J_{\sigma^n(b)}$ is cyclic of order p^n for $n > 0$ (see also the discussion at the bottom of page 3058 in [5]).

So any difference closed field containing K must contain a solution of $\psi(x)$. However, Example 6.7 of [5] shows that if b is as above, and L is any finite extension of $K(b)_\sigma$, then there are $2^{\aleph_0}$ nonisomorphic ways of extending σ to L^{alg}. Thus K does not have a difference closure.

One can build other examples along the same lines, using moduli spaces of abelian varieties.

Example 1.4 (an example in characteristic $p > 0$). Let $K = k(A)_\sigma^{\mathrm{alg}}$, where k is a countable pseudofinite field fixed by σ, σ restricts to a generator of $\mathcal{G}al(k^{\mathrm{alg}}/k)$, and A is the set of solutions of the equation $\sigma(x)^p - \sigma(x) + x^p = 0$ (in some countable difference closed overfield). Then in any difference closed field containing K, the set B of solutions of the equation $\sigma(x) - x^p + x = 0$ is an infinite-dimensional $\mathbb{F}_p$-vector space. However, as was shown in Example 6.5 of [5], there are $2^{|A|}$ ways of extending σ from $Kk(B)_\sigma^{\mathrm{alg}}$ to $K(B)_\sigma^{\mathrm{alg}}$: there is a definable nondegenerate bilinear map $q : A \times B \to \mathbb{F}_p$, which can be chosen totally arbitrarily.

In fact this example is part of a large family of examples: let f and g be additive polynomials with coefficients in a difference field K, and assume that the subgroup A of $\mathbb{G}_a$ defined by $f(x) = g(\sigma(x))$ is locally modular. Then there is a definable subgroup B of $\mathbb{G}_a$, and a definable nondegenerate bilinear map $A \times B \to \mathbb{F}_p$. As above, there is no prime model over $K(A)_\sigma$.

While we provided examples of difference fields not having a difference closure, we did not provide a procedure which, given a difference field which is not difference closed, exhibits a nonisolated type which needs to be realised. So, the following remains open:

Question 1.5. Are there any difference fields which are not difference closed but admit a difference closure?

Omar León Sánchez and Marcus Tressl introduced in [12] the notion of *large* differential fields of characteristic 0, and they showed that their (field-theoretic) algebraic closure are differentially closed, thus showing that the theory DCF_0 can have minimal prime models. One may try introducing the notion of *large* difference field, and hope for a similar result.

2. Preliminaries

Basic difference algebra.

2.1. Let $K \subset \mathcal{U}$ be difference fields. If $X = (X_1, \dots, X_n)$, the ring

$$K[X]_\sigma = K[\sigma^i(X_j)]_{1\le j\le n, i\in\mathbb{N}}$$

is called the *n-fold difference polynomial ring*. A *difference equation* is an equation of the form $f(X) = 0$ for some $f(X) \in K[X]_\sigma$.

If a is a finite tuple in $\mathcal{U}$, and L is a difference subfield of $K(a)_\sigma$ containing K, then $L = K(b)_\sigma$ for some finite tuple b [7, 5.23.18].

An element $a \in \mathcal{U}$ is *transformally algebraic* over K if it satisfies some nontrivial difference equation with parameters in K. Otherwise, it is *transformally transcendental* over K. A tuple a is *transformally algebraic over* K if all its elements are. A (maybe infinite) tuple of elements of $\mathcal{U}$ is transformally independent over K if it does not satisfy any nontrivial difference equation with coefficients in K. A *transformal transcendence basis of* $\mathcal{U}$ *over* K is a subset B of $\mathcal{U}$ which is transformally independent over K and maximal such; every element of K will then be transformally algebraic over $K(B)_\sigma$. We denote by $\Delta(K)$ the *transformal transcendence degree of* K, i.e., the cardinality of a transformal transcendence basis of K, and if L is a difference field containing K, by $\Delta(L/K)$ the cardinality of a transformal transcendence basis of L over K.

2.2. ***The fixed field.*** The fixed field of $\mathcal{U}$ is the field $\mathrm{Fix}(\sigma)(\mathcal{U}) := \{a \in \mathcal{U} \mid \sigma(a) = a\}$. Then $\mathrm{Fix}(\sigma)(\mathcal{U})$ and K are linearly disjoint over their intersection. (Choose n minimal such that there are $c_1, c_2, \dots, c_n \in \mathrm{Fix}(\sigma)$ and $d_1 = 1, d_2, \dots, d_n \in K$

such that $\sum_i c_i d_i = 0$; applying σ we get that $\sum c_i \sigma(d_i) = 0$, and by minimality of n, that $\sigma(d_i) = d_i$ for all i.) This implies in particular that if E is a difference subfield of K, then $E\,\mathrm{Fix}(\sigma)(\mathcal{U})$ and K are linearly disjoint over their intersection $E(\mathrm{Fix}(\sigma)(\mathcal{U}) \cap K)$. In positive characteristic, similar results hold for the other fixed fields $\mathrm{Fix}(\sigma^n\,\mathrm{Frob}^m)$.

Basic model-theoretic facts.

2.3. For references see [5]. The theory ACFA is supersimple, of SU-rank ω. It eliminates imaginaries, but does not eliminate quantifiers. The completions of ACFA are given by describing the isomorphism type of the automorphism σ of the algebraic closure of the prime field [5, Corollary 1.4].

We let $\mathcal{U}$ be a sufficiently saturated model of ACFA, and K a difference subfield of $\mathcal{U}$.

2.4. ***Types, algebraic closure, independence.*** If a is a tuple of elements of $\mathcal{U}$, then $\mathrm{tp}(a/K)$ is determined by the isomorphism type of the difference field $\mathrm{acl}(Ka) = K(a)_\sigma^{\mathrm{alg}}$ over K: a and b have the same type over K if and only if there is a K-isomorphism of difference fields $K(a)_\sigma \to K(b)_\sigma$ which sends a to b and extends to the algebraic closure of $K(a)_\sigma$ [5, Corollary 1.5]. The SU-rank of a over K, denoted by $\mathrm{SU}(a/K)$, is bounded by $\mathrm{tr.deg}(K(a)_\sigma/K)$, and is finite if and only if $\mathrm{tr.deg}(K(a)_\sigma/K)$ is finite (if and only if a is transformally algebraic over K).

Let A, B, C be subsets of $\mathcal{U}$. Then A is *independent from B over C*, denoted $A \mathop{\smile\!\!\!\!\!\!|}_C B$, if and only if the fields $\mathrm{acl}(AC)$ and $\mathrm{acl}(BC)$ are free over $\mathrm{acl}(C)$. Equivalently, if whenever a is a tuple of elements in A, then the prime σ-ideal $I_\sigma(a/\mathrm{acl}(BC)) := \{f(X) \in \mathrm{acl}(BC)[X]_\sigma \mid f(a) = 0\}$ is generated (as a σ-ideal) by its intersection with $\mathrm{acl}(C)[X]_\sigma$. Then independence coincides with nonforking, and we also say, in that case, that $\mathrm{tp}(A/BC)$ *does not fork over C*.

2.5. ***Reducts.*** For an integer $n > 0$, denote by $\mathcal{L}[n]$ the language $\{+, -, \cdot, 0, 1, \sigma^n\}$, and by $\mathcal{U}[n]$ the reduct $(\mathcal{U}, \sigma^n)$ to the language $\mathcal{L}[n]$. By [5, Corollaries to (1.12)], $\mathcal{U}[n] \models$ ACFA. If a is a tuple in $\mathcal{U}$, then $\mathrm{tp}(a/K)[n]$ denotes the type of a in the reduct $\mathcal{U}[n]$, and $\mathrm{qftp}(a/K)[n]$ the quantifier-free type of a in the reduct $\mathcal{U}[n]$.

2.6. ***Notions of canonical bases.*** If a is a tuple in $\mathcal{U}$, then $\mathrm{Cb}(a/K)$ denotes the smallest difference field over which $I_\sigma(a/K)$ is defined. Then $\mathrm{tp}(a/K)$ does not fork over $\mathrm{Cb}(a/K)$. Also, $\mathrm{Cb}(a/K)$ is contained in the algebraic closure over K of finitely many independent realisations of $\mathrm{tp}(a/K)$; if $K(a)_\sigma$ is a regular extension of K, then $\mathrm{Cb}(a/K)$ is contained in the difference field generated over K by finitely many independent realisations of $\mathrm{tp}(a/K)$ (see the proof of Lemma 2.13(4) in [5]). $\overline{\mathrm{Cb}}(a/K)$ denotes $\mathrm{Cb}(a/K)^{\mathrm{alg}}$. Note that a (finitary) type does not fork over some finite set.

2.7. ***The generic type.*** The *generic* 1-type is the type of a transformally transcendental element. It is axiomatised by its quantifier-free part, is definable and stationary.[1] Similarly, if V is a variety defined over the algebraically closed difference field K, then the generic type of V (which is characterised by having a realisation a with $\Delta(K(a)_\sigma/K) = \dim(V)$) is axiomatised by its quantifier-free part, is definable and stationary [5, Corollaries 2.11].

2.8. ***Orthogonality of types.*** Let p and q be (partial) types over A and B, respectively. If $A = B$, we say that p and q are *almost orthogonal* (or *weakly orthogonal*), denoted by $p \perp^a q$, if whenever a realises p and b realises q, then $a \mathop{\smile\!\!\!\!\!\!|}_A b$. We say that p and q are *orthogonal*, denoted by $p \perp q$, if whenever C contains $A \cup B$, and a realises p, b realises q, and $a \mathop{\smile\!\!\!\!\!\!|}_A C$, $b \mathop{\smile\!\!\!\!\!\!|}_B C$, then $a \mathop{\smile\!\!\!\!\!\!|}_C b$.

2.9. ***The dichotomy in characteristic* 0.** Recall that a partial type π over a set A is called *one-based*[2] if whenever $a_1, \dots, a_n$ realise π and $B \supset A$, then $(a_1 \dots a_n) \mathop{\smile\!\!\!\!\!\!|}_C B$, where $C = \operatorname{acl}(Aa_1, \dots, a_n) \cap \operatorname{acl}(B)$.[3]

Types of finite SU-rank are analysable in terms of types of SU-rank 1. The main result of [5] says that in characteristic 0, a type q of SU-rank 1 is either one-based, or nonorthogonal to the fixed field. Moreover, if q is one-based, then it is stable stably embedded and definable. See Theorem 4.10 in [5].

2.10. ***Stable embeddability of the fixed field.*** Recall that a subset S of $\mathcal{U}^n$, which is definable or ∞-definable, is *stably embedded* if whenever $D \subset \mathcal{U}^{nm}$ is definable with parameters from $\mathcal{U}$, then $D \cap S^m$ is definable with parameters from S. An important result of [5] (Proposition 1.11) says that the fixed field $F := \operatorname{Fix}(\sigma)$ of $\mathcal{U}$ is stably embedded: if $D \subset F^n$ is definable in the difference field $\mathcal{U}$ (with parameters from $\mathcal{U}$), then it is definable in the pure field language in F (with parameters from F). In fact, one has more: let $C = \operatorname{acl}(C) \subset \mathcal{U}$, and b a tuple in F. Then $\operatorname{tp}_F(b/C \cap F) \vdash \operatorname{tp}_{\mathcal{U}}(b/C)$; indeed, all finite σ-stable extensions of CF are contained in CF^{alg} (see Lemma 4.2 in [4]), and therefore any $(C \cap F)$-automorphism of the field F extends to a C-automorphism of the difference field $\operatorname{acl}(CF)$, since it obviously extends to a C-automorphism of CF, and the automorphism σ of CF^{alg} extends uniquely to $\operatorname{acl}(CF)$ up to isomorphism over CF^{alg} by Babbitt's theorem (see, e.g., Lemma 2.8 in [5]).

For more properties of stably embedded sets or types, see the appendix of [5].

2.11. ***More on stable stably embedded types.*** For a definition of a (partial) type being *stable stably embedded*, see Lemma 2 of the appendix of [5]. Here we use

[1] A type p over a set A is stationary if whenever $B \supset A$, then p has a unique nonforking extension to B.

[2] In [5], they are called *modular*.

[3] Here we are using the fact that any completion of ACFA eliminates imaginaries.

the following consequence: let $A = \mathrm{acl}(A)$ be algebraically closed, and suppose that $\mathrm{tp}(a/A)$ is stable stably embedded. Then $\mathrm{tp}(a/A)$ is definable (over A; see Lemma 1 in the Appendix of [5]). Also, if $B = \overline{\mathrm{Cb}}(a/A)$ and $\mathrm{tp}(a/B) \perp^a \mathrm{tp}(A/B)$, then $\mathrm{tp}(a/B) \vdash \mathrm{tp}(a/A)$; this is because $\mathrm{tp}(a/B)$ has a unique nonforking extension to any superset of the algebraically closed set B.

Definition 2.12 (internality to the fixed field). Let π be a partial type over $A \subset \mathcal{U}$, and $F = \mathrm{Fix}(\sigma)(\mathcal{U})$.

(1) π is *qf-internal to* $\mathrm{Fix}(\sigma)$ if there is some finitely generated over A difference field C such that whenever a realises π, there is a tuple b in F such that $a \in C(b)$. I.e., $a \in CF$.

(2) π is *almost internal to* $\mathrm{Fix}(\sigma)$ if there is some finitely generated over A difference field C such that whenever a realises π, there is a tuple b in F such that $a \in \mathrm{acl}(Cb)$.

Remarks 2.13. Clearly qf-internality implies almost internality. Moreover, to show qf-internality or almost internality of a (complete) type p, it is enough to do it for a particular realisation a of the type p, i.e., to find C independent from a over A such that $a \in CF$ or $a \in \mathrm{acl}(CF)$. See Lemma 5.2 in [5].

Internality or almost internality (to $\mathrm{Fix}(\sigma)$) of a type is in fact a property of its quantifier-free part.

Recall that a difference field E is linearly disjoint from F over $F \cap E$. It follows that in (1) above, the tuple b can be taken so that $C(b) = C(a)_\sigma$: take a generating tuple d of the (pure) field extension $F \cap C(a)_\sigma$ of $F \cap C$; as F is linearly disjoint from $C(a)_\sigma$ over $F \cap C(a)_\sigma$, we get that CF is linearly disjoint from $C(a)_\sigma$ over $C(d)$, i.e., that $C(a)_\sigma = C(d)$ since $a \in CF$.

Lemma 2.14. *Let* $A = \mathrm{acl}(A)$, *and assume that* $\mathrm{tp}(a/A)$ *is almost internal to* $\mathrm{Fix}(\sigma)$. *Then there is* $a' \in A(a)_\sigma$ *such that* $\mathrm{tp}(a'/A)$ *is qf-internal to* $\mathrm{Fix}(\sigma)$, $\sigma(a') \in A(a')$, *and* $a \in \mathrm{acl}(Aa')$.

Proof. By assumption there is some tuple c independent from a over A and such that $a \in \mathrm{acl}(AFc)$. Taking b in F such that $A(c, a)_\sigma \cap F = (F \cap A)(b)$, we obtain that F is linearly disjoint from $A(c, a)_\sigma$ over $(F \cap A)(b)$, and therefore that $AF(c, b)_\sigma$ and $A(c, a)_\sigma$ are linearly disjoint over $A(c, b)_\sigma$, so that $a \in \mathrm{acl}(Acb)$ (since $a \in \mathrm{acl}(AFcb)$). As c is independent from a over $A = \mathrm{acl}(A)$, it follows that $A(c, a)_\sigma = A(c, a, b)_\sigma$ is a regular extension of $A(a)_\sigma$, and therefore that $\mathrm{Cb}(b, c/A(a)_\sigma)$ is contained in the difference field generated by finitely many independent realisations of $\mathrm{tp}(b, c/A(a)_\sigma)$ (see 2.6). Again, as c is independent from a over A and b is in F, it follows that if a' is such that $\mathrm{Cb}(c, b/A(a)_\sigma) = A(a')_\sigma$, then $\mathrm{tp}(a'/A)$ is qf-internal to $\mathrm{Fix}(\sigma)$. As $b \in A(a', c)_\sigma$ and c is independent from a over A, it follows that $a \in \mathrm{acl}(Aa')$ as desired. As $A(c, a')_\sigma = A(c, b)_\sigma$

and $b \in F$, it follows that $A(c, a')_\sigma$ is finitely generated as a field extension of $A(c)_\sigma$. But as a' and c are independent over A, the same holds of the field extension $A(a')_\sigma/A$, i.e., for some n, $\sigma^n(a') \in A(a', \sigma(a'), \ldots, \sigma^{n-1}(a'))$. We then replace a' by $(a', \sigma(a'), \ldots, \sigma^{n-1}(a'))$. □

2.15. ***The semiminimal analysis.*** Let a be a tuple which is transformally algebraic over K. Thus $\mathrm{SU}(a/K) < \omega$. As $\mathrm{Th}(\mathcal{U})$ is supersimple, there is a sequence $a_1, \ldots, a_n \in \mathrm{acl}(Ka)$ such that $a \in \mathrm{acl}(Ka_1, \ldots, a_n)$, and for every $0 < i \leq n$, $\mathrm{tp}(a_i/\,\mathrm{acl}(Ka_1, \ldots, a_{i-1}))$ is either one-based of rank 1, or almost internal to a non-one-based type of rank 1. This is a classical result in supersimple theories; for a proof in our case in characteristic 0, see Theorem 5.5 in [5]. Note that in characteristic 0, by the dichotomy of 2.9, all non-one-based types of rank 1 are nonorthogonal to $\sigma(x) = x$, and by Lemma 2.14, almost internality to $\mathrm{Fix}(\sigma)$ may be replaced by qf-internality to $\mathrm{Fix}(\sigma)$.

Definition 2.16. Let T be a completion of ACFA, M a model of T.

(1) We say that M is *$\aleph_\varepsilon$-saturated* if whenever $A \subseteq M$ is finite, then every strong 1-type over A is realised in M. Equivalently, as our theory eliminates imaginaries, if every 1-type over $\mathrm{acl}(A)$ is realised in M.

(2) Let κ be an infinite cardinal or $\aleph_\varepsilon$, and $A \subseteq M$. We say that M is *κ-prime over A* if M is κ-saturated, and A-embeds elementarily into every κ-saturated model of $\mathrm{Th}(M, a)_{a\in A}$. When $\kappa = \aleph_\varepsilon$, one also says that M is *a-prime over A*.

(3) Let κ be an infinite cardinal or $\aleph_\varepsilon$. We say that $A \subseteq M$ is *small* if $A = \mathrm{acl}(A_0)$, where A_0 is finite if we are dealing with $\aleph_\varepsilon$-saturation, and has cardinality $< \kappa$ if we are dealing with κ-saturation. We also say that $A \subseteq M$ is *very small* if $A = \mathrm{acl}(A_0)$, where A_0 is finite. Note that a (very) small set is in particular algebraically closed.

(4) Let κ be an infinite cardinal or $\aleph_\varepsilon$, and $A \subseteq M$. A type p over A is *κ-isolated* if it is implied by its restriction to some small subset of $\mathrm{acl}(A)$.

(5) We say that M is *κ-atomic over $A \subseteq M$* if whenever a is a (finite) tuple in M, then $\mathrm{tp}(a/A)$ is κ-isolated. Recall also that M is *atomic over A* if every finite tuple realises an isolated type over A.

(6) We say that $B = \mathrm{acl}(B) \subseteq M$ is *κ-constructed over $A \subseteq M$* if there is a sequence $(d_\alpha)_{\alpha<\mu}$ in $B \setminus A$ such that for every $\alpha < \mu$, $\mathrm{tp}(d_\alpha/\,\mathrm{acl}(Ad_\beta \mid \beta < \alpha))$ is κ-isolated and $B = \mathrm{acl}(Ad_\alpha \mid \alpha < \mu)$.

Remarks 2.17. (1) If κ is a regular cardinal, then κ-atomicity is transitive: if $A \subseteq B \subseteq C \subseteq M$, with B κ-atomic over A and C κ-atomic over B, then C is κ-atomic over A. This is however not necessarily true when κ is singular. However, this holds if $B = \mathrm{acl}(Ab)$ for some finite tuple b (since every finite tuple in B

realises an isolated type over Ab), or if C is atomic over B. (There are stronger statements involving cardinals $\lambda < \mathrm{cf}(\kappa)$.)

(2) If M is a κ-saturated model of T containing A and M is κ-constructed over A, then M is κ-prime over A.

(3) The property of being κ-constructed is preserved under towers and unions of chains indexed by ordinals.

2.18. ***Algebraic translation of the model-theoretic notions.*** Let us translate what the notions of saturation mean in our case. We will be dealing with either uncountable cardinals or $\aleph_\varepsilon$. Recall from 2.4 that $\mathrm{tp}(a/A)$ is entirely determined by the isomorphism type over the difference field generated by A of the difference field $\mathrm{acl}(Aa)$. So, for κ an uncountable cardinal, the κ-saturation of a model M of ACFA simply means that every system of $< \kappa$ difference equations with coefficients in M which has a solution in a difference field extending M already has a solution in M. This is what was called κ-closed in the introduction.

The notion of κ-prime over a difference subfield corresponds to being a κ-closure of that difference field.

In the case of $\aleph_\varepsilon$-saturation, the algebraic description is a little more complicated, and is better expressed in terms of embedding problems: Work inside a large model $\mathcal{U}$, and consider a submodel M of $\mathcal{U}$. Then M is $\aleph_\varepsilon$-saturated if whenever a is a finite tuple of elements of M and b an element of $\mathcal{U}$, there is an $\mathrm{acl}(a)$-embedding of $\mathrm{acl}(a, b)$ inside M.

A similar description holds for κ-saturated, with the base set a of cardinality $< \kappa$: a model M of ACFA is κ saturated if whenever $A \subset M$ is small and b is a finite tuple in some difference field $\mathcal{U}$ containing M, then there is an A-embedding of $\mathrm{acl}(Ab)$ into M. Note that $|A|$-many difference equations are necessary to describe the isomorphism type of $\mathrm{acl}(Ab)$ over A.

3. The results

Results of Hrushovski [10] show that if F is a pseudofinite field and $C \subset F$, then $\mathrm{Th}(F, c)_{c \in C}$ eliminates imaginaries if and only if the absolute Galois group of the relative algebraic closure inside F of the field generated by C is isomorphic to $\hat{\mathbb{Z}}$. It may therefore happen that $\mathrm{Th}(F)$ eliminates imaginaries in the ring language, but it may also happen that extra elements are needed, for instance if F contains $\mathbb{Q}^{\mathrm{alg}}$. The following lemma will therefore be useful when dealing with $\aleph_\varepsilon$-saturation.

Lemma 3.1. *Let F be an $\aleph_\varepsilon$-saturated pseudofinite field and a a finite tuple in F. Then there is a finitely generated subfield A of F containing a and such that*

$$G(A^{\mathrm{alg}} \cap F) \simeq \hat{\mathbb{Z}}.$$

Proof. Let k be the relative algebraic closure inside F of the subfield generated by a, and consider $k(t)$, where t is transcendental over k. Let Q_0 be the set consisting of all integers n which are either prime numbers or 4 and such that $G(k)$ does not have a quotient isomorphic to $\mathbb{Z}/n\mathbb{Z}$. If $\mathrm{char}(k) \neq 0$, we let $Q = Q_0 \setminus \{4\}$, and if $\mathrm{char}(k) = 0$, we let $Q = Q_0 \setminus \{2\}$. If Q is empty, then $G(k) \simeq \hat{\mathbb{Z}}$, and we are done. So, we assume that Q is nonempty.

By Proposition 16.3.5 of [9], for each n, $k(t)$ has a Galois extension L_n which is regular over k and with $\mathcal{G}al(L_n/k(t)) = \mathbb{Z}/n\mathbb{Z}$. Let L be the field composite of all L_n, $n \in Q$. Then $\mathcal{G}al(L/k(t)) \simeq \prod_{n\in Q} \mathbb{Z}/n\mathbb{Z}$. Observe that $L \cap k^{\mathrm{alg}} = k$, because all the L_n are regular extensions of k and Galois over $k(t)$ of relatively prime order.

Take a topological generator σ_0 of $\mathcal{G}al(L/k(t))$, and a topological generator σ_1 of $G(k)$. Let $\sigma \in G(k(t))$ extend $(\sigma_0, \sigma_1) \in \mathcal{G}al(Lk^{\mathrm{alg}}/k(t)) \simeq \mathcal{G}al(L/k(t)) \times G(k)$; then the subfield A of $k(t)^{\mathrm{alg}}$ fixed by σ is a regular extension of k, with Galois group isomorphic to $\hat{\mathbb{Z}}$, since its Galois group is procyclic, projects onto $G(k)$, onto all $\mathbb{Z}/p\mathbb{Z}$ with p a prime, and onto $\mathbb{Z}/4\mathbb{Z}$ if $\mathrm{char}(k) = 0$.

By general properties of pseudofinite fields and by $\aleph_\varepsilon$-saturation of F, there is a k-embedding φ of A inside F, in such a way that $\varphi(A)^{\mathrm{alg}} \cap F = \varphi(A)$. This is classical, and follows for instance from Lemma 20.2.2 in [9]. □

Lemma 3.2. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$, and let K be a difference field with* $\mathrm{Fix}(\sigma)(K)$ *pseudofinite and κ-saturated. Then there is a model $\mathcal{U}$ of* ACFA *containing K which is κ-saturated and with* $\mathrm{Fix}(\sigma)(\mathcal{U}) = \mathrm{Fix}(\sigma)(K)$.

Proof. (Compare with Afshordel's result [1].) Let $\mathcal{U}_1$ be a κ-saturated model of ACFA containing K, and let $\mathcal{U} \subseteq \mathcal{U}_1$ be maximal such that

$$F := \mathrm{Fix}(\sigma)(\mathcal{U}) = \mathrm{Fix}(\sigma)(K).$$

We show that $\mathcal{U}$ satisfies our conclusion. First observe that $\mathcal{U}$ is algebraically closed. Let $A = \mathrm{acl}(A) \subset \mathcal{U}$ be small and let $p \in S_1(A)$. Then p is realised in $\mathcal{U}_1$, and we take some $a \in \mathcal{U}_1$ realising p, with $\mathrm{SU}(a/\mathcal{U})$ minimal. Let $B \supset A$ be small such that $a \perp\!\!\!\perp_B \mathcal{U}$, and replace p by $\mathrm{tp}(a/B)$.

If $\mathrm{tp}(a/\mathcal{U}) \perp^a \mathrm{Fix}(\sigma)$, then $\mathcal{U}(a)_\sigma^{\mathrm{alg}}$ has the same fixed field as $\mathcal{U}$: indeed, $\mathcal{U}(a)_\sigma^{\mathrm{alg}}$ and $\mathrm{Fix}(\sigma)(\mathcal{U}_1)$ are linearly disjoint over their intersection, which is contained in $\mathcal{U}$ and therefore in K. So by maximality of $\mathcal{U}$, $a \in \mathcal{U}$.

Assume now that $\mathrm{tp}(a/\mathcal{U}) \not\perp^a \mathrm{Fix}(\sigma)$. Then there is some small $C \subset \mathcal{U}$ containing B, and a realisation a' of $\mathrm{tp}(a/B)$ such that $C(a')_\sigma \cap \mathrm{Fix}(\sigma)(\mathcal{U}_1)$ contains some element b not in $\mathcal{U}$. We may and do assume that $\mathrm{Fix}(\sigma)(C)$ has absolute Galois group isomorphic to $\hat{\mathbb{Z}}$ (by Lemma 3.1). But as F is κ-saturated, $\mathrm{tp}_F(b/C \cap F)$ is realised in F, by some b_1. Then b_1 realises $\mathrm{tp}(b/C)$ (see the first paragraph of 2.10). Thus, by κ-saturation of $\mathcal{U}_1$, there is some $a_1 \in \mathcal{U}_1$ such that $\mathrm{tp}(a_1, b_1/C) = \mathrm{tp}(a', b/C)$.

But then a_1 realises p, and $\mathrm{SU}(a_1/\mathcal{U}) \le \mathrm{SU}(a'/B) - \mathrm{SU}(b/C) < \mathrm{SU}(a/B)$, which gives us the desired contradiction.

So in both cases, p is realised in $\mathcal{U}$. $\square$

Corollary 3.3. *Let κ be as above, and K an algebraically closed difference field with* $\mathrm{Fix}(\sigma)(K)$ *κ-saturated. If $\mathcal{U}$ is a κ-prime model of* ACFA *over K then* $\mathrm{Fix}(\sigma)(\mathcal{U}) = \mathrm{Fix}(\sigma)(K)$.

Lemma 3.4. *Let $\mathcal{U}$ be an $\aleph_\varepsilon$-saturated model of* ACFA *of characteristic* 0, *and let K be an algebraically closed difference subfield of $\mathcal{U}$ which contains $F := \mathrm{Fix}(\sigma)(\mathcal{U})$. Let $a \in \mathcal{U}$ be such that $p = \mathrm{tp}(a/K)$ is qf-internal to* $\mathrm{Fix}(\sigma)$, *$p \perp^a \mathrm{Fix}(\sigma)$, and assume that $\sigma(a) \in K(a)$. Then there are a (very) small $A \subseteq K$ and a tuple $b \in \mathcal{U}$ of realisations of p such that*

(1) *$FA(b)$ contains all realisations (in $\mathcal{U}$) of* $\mathrm{qftp}(a/A)[\ell]$, *for any $\ell \ge 1$;*

(2) *if $b' \in \mathcal{U}$ realises* $\mathrm{qftp}(b/A)[m]$ *for some $m \ge 1$, then $FA(b')$ contains all realisations (in $\mathcal{U}$) of* $\mathrm{qftp}(a/A)[\ell]$ *for $\ell \ge 1$;*

(3) $\mathrm{tp}(a/A) \vdash \mathrm{tp}(a/K)$, *and* $\mathrm{tp}(b/A) \vdash \mathrm{tp}(b/K)$.

Proof. Let $k \subset K$ be small such that $a \mathop{\smile\!\!\!\!\!\mid}_k K$ and $\mathcal{G}al(\mathrm{Fix}(\sigma)(k)^{\mathrm{alg}}/\mathrm{Fix}(\sigma)(k))$ is isomorphic to $\hat{\mathbb{Z}}$. Then $\sigma(a) \in k(a)$ and kF contains $\mathrm{Fix}(\sigma^\ell)(\mathcal{U})$ for all $\ell \ge 1$. By assumption, there is some small B (in $\mathcal{U}$, by $\aleph_\varepsilon$-saturation of $\mathcal{U}$) independent from a over k such that $a \in BF$. Hence, there is a tuple c in $B(a)_\sigma \cap F = B(a) \cap F$ such that $B(a) = B(c)$ (by Remarks 2.13). Let $D = \mathrm{Cb}(a, c/B)$. Then $D(c) = D(a)$, and $D \subset k(c_1, a_1, \dots, c_n, a_n)$ for some independent realisations (c_i, a_i) of $\mathrm{qftp}(c, a/B)$ (in some elementary extension of $\mathcal{U}$). By $\aleph_\varepsilon$-saturation of $\mathcal{U}$, we may assume that $(c_1, a_1, \dots, c_n, a_n)$ is in $\mathcal{U}$, and is independent from (c, a) over D. We let $b = (a_1, \dots, a_n)$, $A = \overline{\mathrm{Cb}}(k, c_1, a_1, \dots, c_n, a_n/K)$; then $D \subset kF(b)$, and A is small. As A contains $c_1, \dots, c_n$ ($\in F \subset K$) and k, we also have $D \subset A(b)$, whence $a \in FA(b)$. Note that $a \mathop{\smile\!\!\!\!\!\mid}_k A$ since $A \subset K$.

If $a' \in \mathcal{U}$ realises $\mathrm{qftp}(a/A(b))$, then the difference fields $D(a)$ and $D(a')$ are isomorphic. Hence there is some $c' \in D(a') \cap F$ such that $D(c') = D(a')$, i.e., $a' \in FA(b)$.

Let a' be an arbitrary realisation of $\mathrm{qftp}(a/A)$, and let b' be a realisation of $\mathrm{qftp}(b/A)$, which is independent from (b, a') over A. By the previous paragraph (as b' consists of n realisations of $\mathrm{qftp}(a/A(b))$) we know that $b' \in FA(b)$. The difference fields $A(b)$ and $A(b')$ are A-isomorphic, and this isomorphism extends to an isomorphism of difference fields $A(b, a) \to A(b', a')$. Hence, $a' \in FA(b') \subseteq FA(b)$, as desired. If a' realises $\mathrm{qftp}(a/A)[\ell]$ and is independent from D over k, then the σ^ℓ-difference fields $D(a')$ and $D(a)$ are isomorphic over D. Let $f(x)$ be the tuple of rational functions over D such that $f(a) = c$; then $\sigma^\ell(f(a')) = f(a')$ and

$D(a') = D(f(a'))$. Hence a' belongs to $FA(b)$. An argument similar to the one given in the first case shows it for arbitrary realisation of $\mathrm{qftp}(a/A)[\ell]$ and shows (1).

Note that we have in fact shown that $FA(b') = FA(b)$, and so the conclusion of (1) also holds for b'. An easy argument allows to remove the assumption that b' is independent from b over A: let b'' realise $\mathrm{qftp}(b/A)$, independent from (b, b') over A; then by the proof of the first part: $FA(b'') = FA(b)$ and $FA(b') = FA(b'')$.

Working in $\mathcal{U}[\ell]$, and noting that if $m \mid \ell$ then the realisations of $\mathrm{qftp}(a/A)[m]$ also realise $\mathrm{qftp}(a/A)[\ell]$, part (1) gives (2).

For the proof of (3), we first show that every realisation b' of $\mathrm{qftp}(b/A)[\ell]$ (in $\mathcal{U}$) is independent from K over A. Indeed, by (2), we know that $FA(b) = FA(b')$, and therefore $FK(b) = FK(b') = K(b) = K(b')$ (as $A, F \subseteq K$). This implies that $\mathrm{tr.deg}(b/K) = \mathrm{tr.deg}(b'/K)$, and therefore that $b' \mathop{\smile\!\!\!\!\!\mid}_A K$. As $\mathcal{U}$ is $\aleph_\varepsilon$-saturated and A is small, this shows that if $d \in K$, then

$$\mathrm{qftp}(b/A)[\ell] \perp^a \mathrm{qftp}(d/A)[\ell].$$

By Proposition 4.9 of [5], if $\mathrm{tp}(b/A) \not\vdash \mathrm{tp}(b/K)$, then there would be some tuple $d \in K$ and integer $\ell \geq 1$ such that $\mathrm{qftp}(d/A)[\ell] \not\perp^a \mathrm{tp}(b/A)[\ell]$. But as we just saw, this is impossible, and this gives us (3). (This is where the characteristic 0 assumption is crucial.) □

Remark 3.5. In the above notation, note that if $\mathcal{U} \prec \mathcal{U}'$ and $F' = \mathrm{Fix}(\sigma)(\mathcal{U}')$, then $F'A(b)$ contains all realisations of $\mathrm{qftp}(a/A)[\ell]$ in $\mathcal{U}'$, for any $\ell \geq 1$.

Lemma 3.6. *Let K, A, b, $\mathcal{U}$ be as in Lemma 3.4, and let L be a difference subfield of $\mathcal{U}$ containing K. Then there is a small A' containing A such that* $\mathrm{tp}(b/A') \vdash \mathrm{tp}(b/L)$. *In particular*, $\mathrm{tp}(a/A') \vdash \mathrm{tp}(a/L)$.

Proof. Let $A' \subset L$ be small, containing A and such that $b \mathop{\smile\!\!\!\!\!\mid}_{A'} L$. Then the proof of (3) works. □

Corollary 3.7. *Let K and $\mathcal{U}$ be as in Lemma 3.4, and p be a type which is almost internal to* $\mathrm{Fix}(\sigma)$. *Then any K-indiscernible sequence (a_i) of realisations of p in $\mathcal{U}$ is finite.*

Proof. Let $(a_i)_{i<\omega}$ be a sequence of realisations of p in $\mathcal{U}$ which is K-indiscernible. Then either $a_0 \in K$, or $\mathrm{tp}(a_0/K)$ is almost orthogonal to $\mathrm{Fix}(\sigma)$ (since K contains $F := \mathrm{Fix}(\sigma)(\mathcal{U})$). By Lemma 2.14 there is $a_0' \in K(a_0)_\sigma$ such that $\sigma(a_0') \in K(a_0')$, $a_0 \in K(a_0')^{\mathrm{alg}}$ and $\mathrm{tp}(a_0'/K)$ is qf-internal to $\mathrm{Fix}(\sigma)$. It suffices to show the result for $p = \mathrm{tp}(a_0'/K)$. Let b be the finite tuple of realisations of $\mathrm{tp}(a_0'/K)$ given by Lemma 3.4. If $n > d = \mathrm{tr.deg}(K(b)/K)$ and $\mathrm{tp}(a_i', a_i/K) = \mathrm{tp}(a_0', a_0/K)$, then we know that $a_n' \in K(a_0', \ldots, a_{d-1}')^{\mathrm{alg}}$ (because $K \supset F$). Hence the sequence is finite. □

Definition 3.8. We call a type p over a set A *acceptable* (in $K \supset A$) if A is the algebraic closure of a finite tuple, and either $\mathrm{SU}(p) = 1$ and p is one-based, or p is qf-internal to $\mathrm{Fix}(\sigma)$, almost orthogonal to $\mathrm{Fix}(\sigma)$, and if b realises p then $\sigma(b) \in A(b)$, $\mathrm{tp}(b/A) \vdash \mathrm{tp}(b/K)$, and the set of realisations of $\mathrm{qftp}(b/A)[\ell]$ for $\ell \geq 1$, in some model $\mathcal{U}$ of ACFA containing K, is contained in $A(b)\,\mathrm{Fix}(\sigma)(\mathcal{U})$.

Notation 3.9. Let p be a one-based type of SU-rank 1 over the very small set A. If $A \subset B \subset C$, we denote by $p|B$ the unique nonforking extension of p to B, and by $\dim_B p(C)$ the cardinality of a maximal B-independent subset of realisations of $p|B$ in C.

Lemma 3.10. *Let p be an acceptable one-based type over the very small A, and let K be an algebraically closed difference field containing A. We work in a sufficiently saturated model $\mathcal{U}$ of* ACFA. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$.*

(1) *If K contains κ many A-independent realisations of p, then the nonforking extension of p to K is not κ-isolated, and conversely.*

(2) *Assume that* $\dim_A p(K) < \kappa$. *One of the following holds*:

  (a) *There is some $n < \omega$ and realisations $a_0, \ldots, a_{n-1}$ of $p|K$ such that* $\dim_A p(\mathrm{acl}(Ka_0, \ldots, a_{n-1})) \geq \kappa > \dim_A p(K)$. *Furthermore, if n is minimal with this property, then* $\mathrm{tp}(a_0, \ldots, a_{n-1}/K)$ *is κ-isolated (but* $p|\,\mathrm{acl}(Ka_0, \ldots, a_{n-1})$ *is not).*

  (b) *If B is a set of K-independent realisations of $p|K$ of size $\lambda < \kappa$, then* $\dim_A\ p(\mathrm{acl}(KB)) < \kappa$.

Proof. (1) If $C = \mathrm{acl}(C) \subset K$ is small, then C contains $< \kappa$ A-independent realisations of p, so that the nonforking extension of p to C is realised in K, and $p|K$ is not κ-isolated. The converse is clear: the nonforking extension of p to K is implied by its restriction to $\mathrm{acl}(A, p(K))$.

(2) Case (a) is clear by (1) and because dim is additive. So, assume that there is no such n, and let B be as in (b), and $(a_i)_{i<\lambda} \subset B$ a maximal sequence of independent over K realisations of p, and assume that $\lambda < \dim_A p(\mathrm{acl}(KB)) = \mu \geq \kappa$. So $\mathrm{acl}(Ka_i \mid i < \lambda)$ contains a set C consisting of μ many A-independent realisations of p. Then for each $c \in C$, there is some finite $I_c \subset \lambda$ such that $c \in \mathrm{acl}(Ka_i \mid i \in I_c)$. As $\lambda < \mu$, some set I_c appears μ times. Thus $\dim_A p(\mathrm{acl}(Ka_i \mid i \in I_c)) = \mu \geq \kappa$, which contradicts our assumption. □

Remark 3.11. Let p be the generic 1-type over K, and κ an infinite cardinal. Then p is κ-isolated if and only if $\Delta(K) < \kappa$. This follows easily from the description and properties of the generic types (see 2.7).

Definition 3.12. Let $K = \mathrm{acl}(K) \subset L = \mathrm{acl}(L) \subset \mathcal{U}$. We say that L is *normal over* K (in $\mathcal{U}$) if whenever a is a tuple in L, then L contains all realisations of $\mathrm{tp}(a/K)$ in $\mathcal{U}$.

Lemma 3.13. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$, let $K \subseteq L$ be algebraically closed difference subfields of $\mathcal{U}$, where $\mathcal{U}$ is κ-saturated, and suppose* $\mathrm{Fix}(\sigma)(\mathcal{U}) \subset K$. *Assume that $\mathcal{U}$ is κ-atomic over K.*

(1) *Let a be a finite tuple in $\mathcal{U}$. Then $\mathcal{U}$ is κ-atomic over* $\mathrm{acl}(Ka)$.

(2) *Let $B \subset \mathcal{U}$ be transformally independent over K, and assume that either $|B| < \kappa$, or that B is a transformal transcendence basis of $\mathcal{U}$ over K. Then $\mathcal{U}$ is κ-atomic over* $\mathrm{acl}(KB)$.

(3) *If L is normal over K then $\mathcal{U}$ is κ-atomic over L.*

Proof. (1) Clearly $\mathcal{U}$ is κ-atomic over Ka, but we want something stronger. Let $b \in \mathcal{U}$ be a finite tuple, and let C be a small subset of K such that $\mathrm{tp}(a, b/C) \vdash \mathrm{tp}(a, b/K)$. Note that if b' realises $\mathrm{tp}(b/Ca)$ then $b' \mathop{\smile\!\!\!\!\!\mid}_{Ca} K$, since $(a, b') \mathop{\smile\!\!\!\!\!\mid}_C K$ by κ-isolation of $\mathrm{tp}(a, b/K)$. Let us first show the result when $\mathrm{SU}(b/Ca) < \omega$. If $\mathrm{SU}(b/Ca) = 0$, then $b \in \mathrm{acl}(Ca)$ and the result is obvious. The proof is by induction on $\mathrm{SU}(b/Ca)$; using the semiminimal analysis of $\mathrm{tp}(b/\mathrm{acl}(Ca))$ and induction, we may assume that $\mathrm{tp}(b/\mathrm{acl}(Ca))$ is either 1-based of SU-rank 1, or almost internal to $\mathrm{Fix}(\sigma)$. If $\mathrm{tp}(b/\mathrm{acl}(Ca))$ is one-based, then it is also stable, hence has a unique nonforking extension to any superset of $\mathrm{acl}(Ca)$, in particular to $\mathrm{acl}(Ka)$, and by the remark in the previous paragraph, we get the result: $\mathrm{tp}(b/\mathrm{acl}(Ca)) \vdash \mathrm{tp}(b/\mathrm{acl}(Ka))$.

Assume now that $\mathrm{tp}(b/\mathrm{acl}(Ca))$ is almost internal to $\mathrm{Fix}(\sigma)$, and let $b' \in \mathrm{acl}(Cab)$ be such that $b \in \mathrm{acl}(Cab')$, $\sigma(b') \in \mathrm{acl}(Ca)(b')$, and $\mathrm{tp}(b'/\mathrm{acl}(Ca))$ is qf-internal to $\mathrm{Fix}(\sigma)$ (see Lemma 2.14). By Lemma 3.4, there is a finite tuple $e \in \mathrm{acl}(Ka)$ such that $\mathrm{tp}(b'/\mathrm{acl}(e)) \vdash \mathrm{tp}(b'/\mathrm{acl}(Ka))$. Then $\mathrm{tp}(b'/\mathrm{acl}(Cae)) \vdash \mathrm{tp}(b'/\mathrm{acl}(Ka))$, and because $b \in \mathrm{acl}(Cab')$, we get the desired conclusion.

For the general case, because b is a finite tuple, there is a finite tuple $d \in \mathrm{acl}(Cab)$ such that $\mathrm{SU}(d/Ca) < \omega$, and $\mathrm{tp}(b/\mathrm{acl}(Cad))$ is orthogonal to all types of finite SU-rank. (Indeed, this follows from supersimplicity: if $\mathrm{tp}(b/\mathrm{acl}(Ca))$ is nonorthogonal to some type q of finite SU-rank, then there is $b_1 \in \mathrm{acl}(Cab)$ with $0 < \mathrm{SU}(b_1/\mathrm{acl}(Ca)) < \omega$; repeat the procedure with $\mathrm{tp}(b/\mathrm{acl}(Cab_1))$ until it stops.) By the first case, we know that there is a small $C' \subset \mathrm{acl}(Ka)$ containing C such that $\mathrm{tp}(d/\mathrm{acl}(C'a)) \vdash \mathrm{tp}(d/\mathrm{acl}(Ka))$, and that $\mathrm{acl}(Kad)$ is κ-atomic over $\mathrm{acl}(Ka)$. By Remarks 2.17(1), it suffices to show that $\mathrm{tp}(b/\mathrm{acl}(Kad))$ is κ-isolated. By [6, Theorem 5.3] (see also [6, Appendix B]), $\mathrm{tp}(b/\mathrm{acl}(Cad))$ is stationary. But by the first paragraph of the proof, we know that every realisation of $\mathrm{tp}(b/\mathrm{acl}(Cad))$ is independent from K over $\mathrm{acl}(Cad)$, and this gives the result.

(2) If $B = \varnothing$ there is nothing to prove, so suppose it is not. Then $\Delta(K) < \kappa$ by Remark 3.11. Let a be a finite tuple in $\mathcal{U}$, and let $b \subset B$ be a finite tuple such that $a \mathop{\smile\!\!\!\!\!\mid}_{Kb} B$. Let $c \subset a$ be a transformal transcendence basis of $K(a, b)_\sigma$ over $K(b)_\sigma$ (and therefore also over $K(B)_\sigma$). If $c \neq \varnothing$, then $|B| < \kappa$,

$\Delta(K(B)) < \kappa$, and therefore $\mathrm{tp}(c/\mathrm{acl}(KB))$ is κ-isolated. Moreover, as a is transformally algebraic over $K(b, c)_\sigma$, and $B \setminus \{b\}$ is purely transformally transcendental over $K(b, c)_\sigma$, $\mathrm{tp}(a/\mathrm{acl}(Kbc))$ and $\mathrm{tp}(B/\mathrm{acl}(Kbc))$ are orthogonal, and by stationarity of $\mathrm{tp}(B/\mathrm{acl}(Kbc))$, we get that $\mathrm{tp}(B/\mathrm{acl}(Kbc)) \vdash \mathrm{tp}(B/\mathrm{acl}(Kba))$. By symmetry,

$$\mathrm{tp}(a/\mathrm{acl}(Kbc)) \vdash \mathrm{tp}(a/\mathrm{acl}(KBc)).$$

But $\mathrm{tp}(a, b, c/K)$ is κ-isolated, hence so is $\mathrm{tp}(a/\mathrm{acl}(Kbc))$ by (1), and this gives the result.

(3) Let a be a finite tuple in $\mathcal{U}$, and consider $\mathrm{tp}(a/L)$. Let $d \subset a$ be maximal transformally independent over L. If $d \neq \varnothing$, then d is transformally independent over K, which implies that $\Delta(L/K) = 0$ (by normality of L/K), and that $\Delta(K) = \Delta(L) < \kappa$ (by κ-isolation of $\mathrm{tp}(d/K)$). Therefore $\mathrm{tp}(d/L)$ is κ-isolated.

If $\Delta(L/K) \neq 0$, note that by normality of L over K in $\mathcal{U}$, every element of the tuple a which is not in L is transformally algebraic over K. So, replacing a by $a \setminus L$, we may assume they are all transformally algebraic over K, i.e., that $\mathrm{SU}(a/K) < \omega$. We then let $d = \varnothing$.

In both cases, by (2), $\mathcal{U}$ is κ-atomic over $\mathrm{acl}(Kd)$, and the normality of L over K implies the normality of $\mathrm{acl}(Ld)$ over $\mathrm{acl}(Kd)$. Working over $\mathrm{acl}(Kd)$, we may therefore assume that a and $D := \overline{\mathrm{Cb}}(a/L)$ are transformally algebraic over K.

We use induction on $\mathrm{SU}(a/L)$, and using the semiminimal analysis, we find $b \in \mathrm{acl}(Da)$ such that $\mathrm{tp}(a/\mathrm{acl}(Db))$ is either one-based of SU-rank 1, or almost internal to $\mathrm{Fix}(\sigma)$.

If $\mathrm{tp}(a/\mathrm{acl}(Db))$ is almost internal to $\mathrm{Fix}(\sigma)$, then so is $\mathrm{tp}(a/\mathrm{acl}(Lb))$. By Lemma 2.14, there is $a' \in \mathrm{acl}(Lba)$ such that $a \in \mathrm{acl}(Lba')$ and $\mathrm{tp}(a'/\mathrm{acl}(Lb))$ is qf-internal to $\mathrm{Fix}(\sigma)$. By Lemma 3.4, there is a very small $D' \supseteq D$ such that $\mathrm{tp}(a'/\mathrm{acl}(D'b)) \vdash \mathrm{tp}(a'/\mathrm{acl}(Lb))$, and we may choose it so that $a \in \mathrm{acl}(D'ba')$. This shows that $\mathrm{tp}(a/\mathrm{acl}(Lb))$ is κ-isolated, and therefore so is $\mathrm{tp}(a/L)$.

So assume that $p := \mathrm{tp}(a/\mathrm{acl}(Db))$ is one-based of SU-rank 1, and let c be a tuple containing b such that $\mathrm{acl}(Db) = \mathrm{acl}(c) =: C$. We need to show that $\dim_C p(\mathrm{acl}(Lc)) < \kappa$. As $\mathcal{U}$ is κ-atomic over K, we know that $\mathrm{tp}(a, c/K)$ is κ-isolated, and therefore $\dim_C p(\mathrm{acl}(Kc)) < \kappa$. So, if $\dim_C p(\mathrm{acl}(Lc)) \geq \kappa$, then there is some $a' \in \mathrm{acl}(Lc) \setminus \mathrm{acl}(Kc)$ realising p. Recall that by our earlier step, c, a' are transformally algebraic over K, and therefore so is $e = \mathrm{Cb}(Kca'/L)$. Consider now $\mathrm{acl}(Kca') \cap \mathrm{acl}(Ke) =: E \subset L$; by Proposition 3.1 of [2], $\mathrm{tp}(e/E)$ is almost internal to $\mathrm{Fix}(\sigma)$, and therefore orthogonal to all one-based types. As $\mathrm{tp}(a'/Kc)$ is one-based, and $a' \in \mathrm{acl}(Kce) \setminus \mathrm{acl}(Kc)$, it follows that $e \in E$, since almost internality to $\mathrm{Fix}(\sigma)$ and nonorthogonality to a one-based type imply algebraicity. That is, $e \in \mathrm{acl}(Kca') \cap L$, and as $a' \notin \mathrm{acl}(Kc)$, the tuples a' and e are equialgebraic over $\mathrm{acl}(Kc)$. Hence $\mathrm{acl}(Kca)$ contains a realisation of $\mathrm{tp}(e/\mathrm{acl}(Kc))$, because

$\mathrm{tp}(a/\mathrm{acl}(Kc)) = \mathrm{tp}(a'/\mathrm{acl}(Kc))$. But this contradicts the normality of $\mathrm{acl}(Lc)$ over $\mathrm{acl}(Kc)$. So, $\dim_C(p(\mathrm{acl}(Lc)) < \kappa$, and $\mathrm{tp}(a/Lb)$ is κ-isolated. □

Theorem 3.14. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$, and let K be an algebraically closed difference field of characteristic* 0 *such that $F := \mathrm{Fix}(\sigma)(K)$ is pseudofinite and κ-saturated.*

(1) *Then there is a κ-prime model $\mathcal{U}$ over K.*

(2) *Furthermore, $\mathcal{U}$ is κ-atomic over K, and every sequence of K-indiscernibles has length $\leq \kappa$ (i.e., if $\kappa = \aleph_\varepsilon$, $\leq \aleph_0$; by convention, if κ is meant as a cardinal, then $\aleph_\varepsilon$ will mean $\aleph_0$).*

Proof. By Lemma 3.2, there is a κ-saturated model $\mathcal{U}_1$ of ACFA containing K and with fixed field $F = \mathrm{Fix}(\sigma)(K)$. We will construct a submodel $\mathcal{U}$ of $\mathcal{U}_1$ which is κ-prime over K and satisfies (2). This $\mathcal{U}$ will be κ-constructed.

Step 0. Taking care of the transformal transcendence degree.

If the transformal transcendence degree of K is $< \kappa$, then as any κ-saturated model of ACFA has transformal transcendence degree at least κ, we enlarge K as follows: let $B \subset \mathcal{U}_1$ be a set which is transformally independent over K and of cardinality κ; by [5, Corollaries 2.11], this condition completely determines the K-isomorphism type of $K(B)_\sigma^{\mathrm{alg}}$, and therefore any κ-prime model contains a K-isomorphic copy of $K(B)_\sigma^{\mathrm{alg}}$. We let $K_0 = K(B)_\sigma^{\mathrm{alg}}$. We need to show (2). Each finite subset of B realises a κ-isolated type over K, since the transformal transcendence degree of K is $< \kappa$. Moreover, every tuple in K_0 realises an isolated type over $K(B)_\sigma$; hence K_0 is κ-atomic over K. It is also κ-constructed over K.

Let $(a_i)_{i<\lambda} \subset K_0$ be a K-indiscernible sequence and λ a cardinal. If the a_i are transformally independent over K, then we know that $|\lambda| \leq \kappa$. If not, then by indiscernibility, the transformal transcendence degree of $K(a_i \mid i < \lambda)_\sigma$ over K is finite, and we choose a finite subset c of B such that $K(a_i \mid i < \lambda)_\sigma$ is transformally algebraic over $K(c)_\sigma$. As the elements of B are transformally independent over K, this implies that all the a_i are in fact algebraic over $K(c)_\sigma$. Consider now $D := \overline{\mathrm{Cb}}(c/Ka_i \mid i < \lambda)$. For every i, we know that $a_i \in K(c)_\sigma^{\mathrm{alg}}$, and therefore by definition of D, $a_i \in D(c)_\sigma^{\mathrm{alg}}$. But c is finite, D is contained in the algebraic closure of a finite set (by 2.6), and therefore $D(c)_\sigma^{\mathrm{alg}}$ is countable. Hence so is λ. This shows condition (2) for the extension K_0/K.

We build a sequence K_n, $n < \omega$, of algebraically closed difference subfields of $\mathcal{U}_1$ such that

(i) if p is an acceptable type over a very small $A \subset K_n$, then K_{n+1} contains κ-many A-independent realisations of p;

(ii) K_{n+1} is κ-constructed over K_n.

We let $K_0 = K$ if the transformal transcendence degree of K is $\geq \kappa$, and $K(B)_\sigma^{\mathrm{alg}}$ as in step 0 otherwise. We assume K_n constructed and wish to build K_{n+1}. Let p_β, $\beta < \lambda$, be an enumeration of the acceptable types in K_n, with corresponding very small bases A_β.

Step 1. Defining $K_{n+1} = \bigcup_{\beta<\lambda} K'_\beta$.

We build the sequence K'_β by induction on β, and let $K'_0 = K_n$. If β is a limit ordinal, then we let $K'_\beta = \bigcup_{\gamma<\beta} K'_\gamma$, and $K_{n+1} = K'_\lambda$. We build them so that $K'_{\beta+1}$ satisfies the following:

(i′) $K'_{\beta+1}$ contains κ-many A_β-independent realisations of p_β;

(ii′) $K'_{\beta+1}$ is κ-constructed over K'_β.

Assume K'_β constructed. If p_β has κ-many A_β-independent realisations in K'_β, then we let $K'_{\beta+1} = K'_\beta$. Otherwise, we need to distinguish two cases:

Case 1: p_β is one-based.

Let a_i, $i < \kappa$, be a sequence of K'_β-independent realisations of p_β (a priori, in some elementary extension of $\mathcal{U}_1$). By Lemma 3.10, either there is $n < \omega$ such that $\mathrm{acl}(K'_\beta, a_i \mid i < n)$ contains κ-many A_β-independent realisations of p_β; in that case, taking a minimal such n, $\mathrm{tp}(a_0, \dots, a_{n-1}/K'_\beta)$ is κ-isolated and therefore realised in $\mathcal{U}_1$, so that we may assume $a_0, \dots, a_{n-1} \in \mathcal{U}_1$ and we set $K'_{\beta+1} = \mathrm{acl}(K'_\beta, a_i \mid i < n)$. Then (i′) and (ii′) follow.

If there is no such n, then for every $\lambda < \kappa$, $\mathrm{acl}(K'_\beta a_i \mid i < \lambda)$ does not contain κ-many A_β-independent realisations of p; by the same reasoning we may assume the a_i are in $\mathcal{U}_1$ and we define $K'_{\beta+1} = \mathrm{acl}(K a_i \mid i < \kappa)$. Then (i′) and (ii′) again are satisfied.

Case 2: Not case 1.

Let $a_\beta \in \mathcal{U}_1$ realise p_β, $K'_{\beta+1} = K'_\beta(a_\beta)^{\mathrm{alg}}$. By assumption on p_β, we have $\mathrm{tp}(a_\beta/A_\beta) \vdash \mathrm{tp}(a_\beta/K_n)$. By Lemma 3.6, there is a very small subset B of K'_β which contains A_β and is such that $\mathrm{tp}(a_\beta/B) \vdash \mathrm{tp}(a/K'_\beta)$. So, $\mathrm{tp}(a_\beta/K'_\beta)$ is κ-isolated. We let $K'_{\beta+1} = K'_\beta(a_\beta)_\sigma^{\mathrm{alg}}$. We know that $FK'_\beta(a_\beta)_\sigma$ contains all realisations of $\mathrm{tp}(a_\beta/B)$ in $\mathcal{U}_1$. But since $\mathcal{U}_1$ is κ-saturated, it therefore contains κ independent realisations of $\mathrm{tp}(a_\beta/A_\beta)$, which shows (i′).

We now define $\mathcal{U} = \bigcup_{n\in\omega} K_n$.

Step 2. Show that $\mathcal{U}$ is κ-saturated.

Let $C \subset \mathcal{U}$ be small, and p a 1-type over C, realised by a in $\mathcal{U}_1$. If $\mathrm{SU}(p) = \omega$, then a is transformally transcendental over C; as C is small, K_0 contains a realisation of p. So we may assume that $\mathrm{SU}(p) < \omega$, and the proof is by induction on $\mathrm{SU}(p)$: we assume that for any small D, any 1-type q over D of smaller SU-rank than p is realised in $\mathcal{U}$.

If $\mathrm{SU}(p) = 0$ there is nothing to prove, as p is realised in C. If there is some $b \in C(a)_\sigma^{\mathrm{alg}}$ such that $0 < \mathrm{SU}(b/C) < \mathrm{SU}(a/C)$, then we get the result by induction: $\mathrm{tp}(b/C)$ is realised by some $b' \in \mathcal{U}$, and there is $a' \in \mathcal{U}$ such that $\mathrm{tp}(a', b'/C) = \mathrm{tp}(a, b/C)$ since $\mathrm{acl}(Cb')$ is small and $\mathrm{SU}(a/Cb) < \mathrm{SU}(p)$.

Hence we may assume that there is no such b, whence p is either one-based of SU-rank 1 or almost internal to $\mathrm{Fix}(\sigma)$ (by the semiminimal analysis of 2.15). We need to distinguish three cases.

Case 1: p is one-based of SU-rank 1.

Let $A \subset C$ be very small such that p does not fork over A. Let $n < \omega$ be such that $A \subset K_n$; then p, being acceptable, occurs as a p_β, and is therefore realised in K_{n+1}.

Case 2: p is realised in $\mathrm{Fix}(\sigma)$.

If $a \in \mathrm{Fix}(\sigma)$, we saw in 2.10 that $\mathrm{tp}_F(a/C \cap F) \vdash \mathrm{tp}(a/C)$. The saturation hypothesis on F then gives the result: p is realised in F.

Case 3: Assume now that $p \perp^a \mathrm{Fix}(\sigma)$, p almost internal to $\mathrm{Fix}(\sigma)$.

By Lemma 2.14, there is $a_1 \in C(a)_\sigma$ such that $\mathrm{tp}(a_1/C)$ is qf-internal to $\mathrm{Fix}(\sigma)$, $\sigma(a_1) \in C(a_1)$, and $a \in C(a_1)^{\mathrm{alg}}$. We may replace p by $\mathrm{tp}(a_1/C)$, i.e., assume that p is qf-internal to $\mathrm{Fix}(\sigma)$. Let $C_0 \subset C$ be very small such that p does not fork over C_0. By Lemma 3.4 there is a tuple b of realisations of p and a very small D containing C_0, contained in $\mathrm{acl}(CF)$, such that $FD(b)$ contains all realisations of $\mathrm{qftp}(a/D)$, and $\mathrm{tp}(b/D) \vdash \mathrm{tp}(b/\mathrm{acl}(CF))$. Thus, $\mathrm{tp}(b/D)$ is acceptable, and if n is such that $D \subset K_n$, then p in realised in K_{n+1}.

Step 3. $\mathcal{U}$ is κ-prime over K.

This is clear, by Remarks 2.17(2)–(3).

Step 4. $\mathcal{U}$ is κ-atomic over K.

When κ is regular or $\aleph_\varepsilon$, then this follows from $\mathcal{U}$ being κ-constructed over K. The proof in the singular case is a little more delicate, and is done by induction. We already saw that K_0 is κ-atomic over K. Let a be a finite tuple in $\mathcal{U}$, and (in the notation of Step 1) choose n minimal such that $a \in K_{n+1}$, and β minimal such that $a \in K'_{\beta+1}$. If $n = -1$, there is nothing to prove (by Step 0), so assume $n \geq 0$. By definition of $K'_{\beta+1}$, there are a tuple b in K'_β and a tuple c of realisations of p_β such that $a \in \mathrm{acl}(Kbc)$. We may assume that $\mathrm{acl}(Kb)$ contains A_β, and that $c \mathop{\smile\!\!\!\!\!\mid}_{Kb} K'_\beta$. By the induction hypothesis, $\mathrm{tp}(b/K)$ is κ-isolated, and it therefore suffices to show that $\mathrm{tp}(c/\mathrm{acl}(Kb))$ is κ-isolated (by Remarks 2.17(1)). If p_β is qf-internal to $\mathrm{Fix}(\sigma)$ then we know by Lemma 3.4 that there is some very small $D \subset \mathrm{acl}(Kb)$ such that $\mathrm{tp}(c/D) \vdash \mathrm{tp}(c/\mathrm{acl}(Kb))$, and we are done.

If p_β is one-based, then we may assume that the elements of the tuple c are independent over K'_β, maybe at the cost of increasing $b \in K'_\beta$. Then, by the construction of $K'_{\beta+1}$ in Step 1, we know that $\mathrm{tp}(c/K'_\beta)$ is κ-isolated, so that if c' is a proper subtuple of c (consisting of realisations of p_β), then $\dim_{A_\beta} p_\beta(\mathrm{acl}(K'_\beta c')) < \kappa$. In particular, $\dim_{A_\beta} p(\mathrm{acl}(Kbc')) < \kappa$, and $\mathrm{tp}(c/\mathrm{acl}(Kb))$ is κ-isolated (by Lemma 3.10).

Remark (notation as in Step 1 and above). The same proof shows that $\mathcal{U}$ is κ-atomic over each K_n, and over each K'_β. Moreover, the fact that $\mathcal{U}$ is κ-atomic over K'_β implies that $p_\beta(\mathcal{U}) \subset K'_{\beta+1}$.

Step 5. If $(b_i)_{i<\lambda} \subset \mathcal{U}$ is K-indiscernible, with λ a cardinal, then $\lambda \leq \kappa$.

By supersimplicity, for some $n < \omega$ the elements b_i, $n < i < \lambda$, are independent over $K(b_0, \dots, b_n)_\sigma$. If $\mathrm{SU}(b_{n+1}/Kb_0, \dots, b_n) \geq \omega$, then the tuple b_{n+1} contains an element which is transformally transcendental over K, and as the transformal transcendence degree of $\mathcal{U}$ over K is $\leq \kappa$, we get $\lambda \leq \kappa$. So we may assume $\mathrm{SU}(b_{n+1}/\mathrm{acl}(Kb_0, \dots, b_n)) < \omega$.

Let $L = \mathrm{acl}(Kb_0, \dots, b_n)$. Then the sequence $(b_i)_{n<i<\lambda}$ is indiscernible over L. Note that the sequence $\mathrm{acl}(Lb_i)$, $n < i < \lambda$, is also indiscernible over L under a suitable enumeration of each $\mathrm{acl}(Lb_i)$. Hence, if $c_{n+1} \in \mathrm{acl}(Lb_{n+1})$, there are $c_i \in \mathrm{acl}(Lb_i)$, $n+1 < i < \lambda$, such that the sequence $(c_i)_{n<i<\lambda}$ is indiscernible over L. Using the semiminimal analysis (2.15) we may therefore assume that either $\mathrm{tp}(c_i/L)$ is one-based of SU-rank 1, or that $\mathrm{tp}(c_i/L)$ is almost internal to $\mathrm{Fix}(\sigma)$. If $\mathrm{tp}(c_i/L)$ is almost internal to $\mathrm{Fix}(\sigma)$, then the result follows by Corollary 3.7. The one-based case is a little more complicated.

Towards a contradiction, assume that $\lambda > \kappa$ and $\mathrm{tp}(c_{n+1}/L)$ is one-based of SU-rank 1, let $C \subset L$ be a very small set such that $\mathrm{tp}(c_{n+1}/L)$ does not fork over C, and set $p = \mathrm{tp}(c_{n+1}/C)$. Then the tuples c_i, $n < i < \lambda$, form a Morley sequence over C and over L. Let N be κ-prime over $M := \mathrm{acl}(L, c_i \mid n < i < \kappa)$. We may assume that $N \prec \mathcal{U}$.

Claim. $\mathcal{U}$ is κ-prime over L.

It suffices to show that $\mathcal{U}$ is κ-constructed over L. To do that it is enough to show that each LK_m is κ-constructed over LK_{m-1}.

If $m = 0$ and $K_0 \neq K$, let B_0 be a finite subset of B (the transformal transcendence basis of $\mathcal{U}$ over K) such that $b := (b_0, \dots, b_n)$ is independent from K_0 over $\mathrm{acl}(KB_0)$. In particular, b is transformally algebraic over $\mathrm{acl}(KB_0)$, and therefore $\mathrm{tp}(B/\mathrm{acl}(KB_0)) \vdash \mathrm{tp}(B/\mathrm{acl}(LB_0))$ (reason as in the proof of Lemma 3.13(1)), and as $\mathrm{tp}(B_0/L)$ is κ-isolated, it follows that K_0 is κ-constructed over L.

Assume now $m > 0$, and that we have shown that LK'_β is κ-constructed over L. If p_β is not one-based, then by Lemma 3.6, $\mathrm{tp}(a_\beta/\mathrm{acl}(LK'_\beta))$ is κ-isolated, and we are done. Assume now that p_β is one-based; by construction there is a set $(a_\alpha)_{\alpha<\mu}$ of K'_β-independent realisations of $p_\beta|K'_\beta$ such that $K'_{\beta+1} = \mathrm{acl}(K'_\beta, a_\alpha \mid \alpha < \mu)$, and either $\mu \in \omega$ or $\mu = \kappa$.

If $\mu \in \omega$, as $\mathcal{U}$ is κ-atomic over K'_β, we get that $\mathrm{tp}(a_0, \dots, a_{\mu-1}, b/K'_\beta)$ is κ-isolated and therefore $LK'_{\beta+1}$ is κ-constructed over LK'_β. If $\mu = \kappa$, then $\dim_{K'_\beta} p_\beta(\mathrm{acl}(K'_\beta, b, a_\gamma \mid \gamma < \alpha)) < \kappa$ for each $\alpha < \kappa$, so that $LK'_{\beta+1}$ is κ-constructed over LK'_β (here we use that $p_\beta(\mathcal{U}) \subset K'_{\beta+1}$ and that b is finite).

Hence, $\mathcal{U}$ being κ-prime over L, there is an L-embedding f of $\mathcal{U}$ into N. So we have $L \subset f(\mathcal{U}) \prec N \prec \mathcal{U}$. As $\lambda > \kappa$ and the c_i are independent over L, there is some $n < j < \lambda$ such that $f(c_j) \notin M$. But $\dim_M(p) \geq \kappa$, and by Lemma 3.10, $p|M$ is not isolated. But N is κ-atomic over M, and $f(c_j)$ realises p and is not in M, which gives us the desired contradiction. This finishes the proof of (2) and of the theorem. □

Proposition 3.15. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$, and let $\mathcal{U}$ and $\mathcal{U}'$ be κ-saturated models of* ACFA *of characteristic* 0. *Assume that $\mathcal{U}$ (resp., $\mathcal{U}'$) contains an algebraically closed difference field K (resp., K'), over which it is κ-atomic and over which every sequence of indiscernibles has length $\leq \kappa$. Assume moreover that $F := \mathrm{Fix}(\sigma)(K) = \mathrm{Fix}(\sigma)(\mathcal{U})$, $\mathrm{Fix}(\sigma)(K') = \mathrm{Fix}(\sigma)(\mathcal{U}')$, and that we have an isomorphism $f : K \to K'$. Let p be an acceptable type over some very small $A \subset K$, and $p' = f(p)$. If $L = \mathrm{acl}(Kp(\mathcal{U}))$ and $L' = \mathrm{acl}(Kp'(\mathcal{U}'))$, then f extends to an isomorphism between L and L'.*

Proof. Note that p' is also acceptable, with very small basis $A' = f(A)$. If p is not one-based, then this is clear by Lemma 3.4: $L = \mathrm{acl}(Kb)$, $L' = \mathrm{acl}(K'b')$ for some tuples b realising p and b' realising p'. We extend $f|A$ to an isomorphism $g_0 : \mathrm{acl}(Ab) \to \mathrm{acl}(A'b')$ which sends b to b'; as $\mathrm{tp}(b/A) \vdash \mathrm{tp}(b/K)$ and $\mathrm{tp}(b'/A') \vdash \mathrm{tp}(b'/K')$, $g_0 \cup f$ extends to an isomorphism $\mathrm{acl}(Kb) \to \mathrm{acl}(K'b')$.

Assume now that p is one-based. Any κ-saturated model of ACFA containing A contains (at least) κ realisations of p which are independent over A; hence so do $\mathcal{U}$ and $\mathcal{U}'$. Let $(a_i)_{i<\lambda} \subset \mathcal{U}$ be a set of realisations of p which is maximal independent over K, with λ a cardinal, and let $(a'_i)_{i<\mu} \subset \mathcal{U}'$ be defined analogously over K'. By Lemma 3.10 and our hypothesis on the length of K-indiscernible sequences, either λ is finite or $\lambda = \kappa$. If $\lambda = n < \omega$, then as $\mathrm{tp}(a'_0, \dots, a'_{n-1}/K') = f(\mathrm{tp}(a_0, \dots, a_{n-1}/K))$, it follows that $\mathrm{acl}(K'a'_0, \dots, a'_{n-1})$ contains κ-many independent realisations of $f(p)$, so that $\mu \leq n$. The symmetric argument gives $\mu = \lambda$. Define g on $K(a_i \mid i < \lambda)_\sigma$ by $g(a_i) = a'_i$, and extend to $L = \mathrm{acl}(Ka_i \mid i < \lambda)$. □

Theorem 3.16. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$, and let $\mathcal{U}$ and $\mathcal{U}'$ be κ-saturated models of* ACFA *of characteristic* 0 *containing an algebraically closed difference field K, with $F := \mathrm{Fix}(\sigma)(K) = \mathrm{Fix}(\sigma)(\mathcal{U}) = \mathrm{Fix}(\sigma)(\mathcal{U}')$. Assume that $\mathcal{U}$ and $\mathcal{U}'$ are κ-atomic over K, and that any sequence of K-indiscernibles in $\mathcal{U}$ or in $\mathcal{U}'$ has length $\leq \kappa$. Then $\mathcal{U} \simeq_K \mathcal{U}'$.*

Proof. We start with the generic type: if the transformal transcendence degree of K is $\geq \kappa$, then $\mathcal{U}$ and $\mathcal{U}'$ are transformally algebraic over K. If not, then let D be a transformal transcendence basis of $\mathcal{U}$ over K and D' a transformal transcendence basis of $\mathcal{U}'$ over K. They have the same cardinality κ, and there is a K-isomorphism $K(D)_\sigma^{\mathrm{alg}} \to K(D')_\sigma^{\mathrm{alg}}$. By Lemma 3.13, $\mathcal{U}$ and $\mathcal{U}'$ still satisfy the hypotheses over $K(D)_\sigma^{\mathrm{alg}}$ and $K(D')_\sigma^{\mathrm{alg}}$. Hence we may assume that both $\mathcal{U}$ and $\mathcal{U}'$ are transformally algebraic over K. We define by induction on n an increasing sequence K_n of algebraically closed subfields of $\mathcal{U}$ such that for each n, if p is an acceptable type over some (very small) $A \subset K_{n-1}$, then K_n contains all realisations of p in $\mathcal{U}$, and furthermore, $K_n = \mathrm{acl}(K_{n-1}P)$ for the set P of all realisations (in $\mathcal{U}$) of acceptable types over some subset of K_{n-1}. Then each K_n is normal over K_{n-1} (and in fact over K), and so by Lemma 3.13, $\mathcal{U}$ satisfies the hypotheses over K_n. Note also that $\mathcal{U} = \bigcup_{n<\omega} K_n$. We let $L_n \subset \mathcal{U}'$ be defined analogously. It then suffices to build a sequence g_n of K-isomorphisms $K_n \to L_n$.

Assume g_{n-1} already built. Let p_β, $\beta < \lambda$, be an enumeration of all acceptable types over a subset of K_{n-1}, with associated small basis A_β. Note that $f(p_\beta)$, $\beta < \lambda$, enumerates all acceptable types over subsets of L_{n-1}, since if q is an acceptable type over the very small $C \subset L_{n-1}$, so is $g_{n-1}^{-1}(q)$ (over $g_{n-1}^{-1}(C) \subset K_{n-1}$). We build by induction on $\beta < \lambda$ an increasing sequence K'_β of algebraically closed difference subfields of $\mathcal{U}$ such that K'_β contains all realisations in $\mathcal{U}$ of p_γ for all $\gamma < \beta$. Assume we have extended g_{n-1} to an isomorphism $f_\beta : K'_\beta \to L'_\beta$, where L'_β contains all realisations in $\mathcal{U}'$ of $g_{n-1}(p_\gamma)$ for all $\gamma < \beta$. As $\mathcal{U}$ is κ-atomic over K_{n-1}, it is also κ-atomic over K'_β (by Lemma 3.13), and similarly, $\mathcal{U}'$ is κ-atomic over $L'_\beta = f_\beta(K'_\beta)$. Extending f_β to an isomorphism $f_{\beta+1} : K'_{\beta+1} \to L'_{\beta+1}$ is given by Proposition 3.15.

As remarked before, if q is an acceptable type over some $A' \subset L'_{n-1}$, then $g_{n-1}^{-1}(q) = p_\beta$ for some $\beta < \lambda$, and so L'_n contains $q(\mathcal{U}')$, and K'_n contains $g_{n-1}^{-1}(q)(\mathcal{U})$. This finishes the induction step. Then $g = \bigcup_{n<\omega} g_n$ is a K-isomorphism between $\mathcal{U}$ and $\mathcal{U}'$. □

Theorem 3.17. *Let κ be an uncountable cardinal or $\aleph_\varepsilon$ and let K be an algebraically closed difference field of characteristic* 0, *with* $\mathrm{Fix}(\sigma)(K)$ *pseudofinite and κ-saturated. Then* ACFA *has a κ-prime model over K, and it is unique up to K-isomorphism.*

Proof. This follows immediately from Theorem 3.16 together with Theorem 3.14, as the properties are preserved by elementary substructures. □

Remark 3.18. Note that the result also holds under the weaker hypothesis that K is algebraically closed, $|\mathrm{Fix}(\sigma)(K)| < \kappa$, and $\kappa^{<\kappa} = \kappa \geq \aleph_1$, so that the theory of pseudofinite fields has a unique (up to K-isomorphism) saturated model of cardinality κ containing $\mathrm{Fix}(\sigma)(K)$. (This uses the stable embeddability of $\mathrm{Fix}(\sigma)$; see 2.10.)

References

[1] B. Afshordel, "Generic automorphisms with prescribed fixed fields", *J. Symb. Log.* **79**:4 (2014), 985–1000. MR Zbl

[2] Z. Chatzidakis, "A note on canonical bases and one-based types in supersimple theories", *Confluentes Math.* **4**:3 (2012), art. id. 1250004. MR Zbl

[3] Z. Chatzidakis, "A note on the non-existence of prime models of theories of pseudo-finite fields", preprint, 2020. arXiv 2004.05593

[4] Z. Chatzidakis, C. Hardouin, and M. F. Singer, "On the definitions of difference Galois groups", pp. 73–109 in *Model theory with applications to algebra and analysis*, vol. I, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **349**, Cambridge Univ. Press, 2008. MR Zbl

[5] Z. Chatzidakis and E. Hrushovski, "Model theory of difference fields", *Trans. Amer. Math. Soc.* **351**:8 (1999), 2997–3071. MR Zbl

[6] Z. Chatzidakis and E. Hrushovski, "Model theory of endomorphisms of separably closed fields", *J. Algebra* **281**:2 (2004), 567–603. MR Zbl

[7] R. M. Cohn, *Difference algebra*, Interscience Publishers, New York, 1965. MR Zbl

[8] J.-L. Duret, "Les corps faiblement algébriquement clos non séparablement clos ont la propriété d'indépendence", pp. 136–162 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

[9] M. D. Fried and M. Jarden, *Field arithmetic*, 3rd ed., Ergebnisse der Math. (3) **11**, Springer, 2008. MR Zbl

[10] E. Hrushovski, "Pseudo-finite fields and related structures", pp. 151–212 in *Model theory and applications*, edited by L. Bélair et al., Quad. Mat. **11**, Aracne, Rome, 2002. MR Zbl

[11] S. Lang, *Elliptic functions*, 2nd ed., Grad. Texts in Math. **112**, Springer, 1987. MR Zbl

[12] O. León Sánchez and M. Tressl, "Differentially large fields", preprint, 2020. To appear in *Algebra & Number Theory*. arXiv 2003.02886

[13] A. Pillay and S. Shelah, "Classification theory over a predicate, I", *Notre Dame J. Formal Logic* **26**:4 (1985), 361–376. MR Zbl

[14] S. Shelah and A. Usvyatsov, "Classification over a predicate — the general case: Part I — structure theory", preprint, 2019. arXiv 1910.10811

[15] J. H. Silverman, *The arithmetic of elliptic curves*, 2nd ed., Grad. Texts in Math. **106**, Springer, 2016. MR Zbl

Received 29 Aug 2022. Revised 31 Dec 2022.

ZOÉ CHATZIDAKIS:

zoe.chatzidakis@imj-prg.fr
Université Paris Cité and Sorbonne Université, CNRS, IMJ-PRG, F-75013 Paris, France

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.429

An exposition of Jordan's original proof of his theorem on finite subgroups of $\mathrm{GL}_n(\mathbb{C})$

Emmanuel Breuillard

Dedicated to Udi Hrushovski on the occasion of his 60th birthday.

We discuss Jordan's theorem on finite subgroups of invertible matrices and give an account of his original proof.

1. Introduction

In 1878 Camille Jordan [29] proved the following theorem:

Theorem 1.1 (Jordan's theorem). *Let G be a finite subgroup of $\mathrm{GL}_n(\mathbb{C})$. Then there is a normal abelian subgroup A in G of index bounded by a constant $J(n)$ depending on n only.*

It is the purpose of this note to provide an account of Jordan's original proof of his result. Jordan's proof is purely algebraic, and quite different from the proofs found in most textbooks (such as [19] or [21]) that are based on a geometric argument due to Bieberbach [3]. Jordan's proof does not appear to have been discussed much elsewhere (with the exception of Dieudonné's notes in Jordan's collected works [20]) even as this year marks the hundredth anniversary of Jordan's death.

Jordan's motivation for proving this result came from the study of linear differential equations of order n with rational functions as coefficients and with algebraic solutions; in this context finite subgroups of GL_n arise naturally as monodromy groups and information such as Theorem 1.1 on the monodromy group translates immediately into structural properties for the solutions of the equation.[1] Prior to Jordan, Fuchs and Klein had studied the two dimensional case and Klein had given a complete list of finite subgroups of $\mathrm{GL}_2(\mathbb{C})$. Jordan announced his result in [28], published it in [29] and later wrote a second article [30] to clarify his proof.

Jordan argued by induction on the dimension, but he gave no explicit bound on $J(n)$ in his article, not even an inductive one. It is therefore understandable

MSC2020: 20C10, 20E34.
Keywords: Jordan's theorem, finite subgroups of matrices, nonstandard analysis.

[1] For the full story of the motivations and context in which Jordan's theorem was proven, we refer the reader to the wonderful book by Jeremy Gray [24].

that mathematicians sought to find explicit bounds closer to the truth and this topic has been quite active in the last 146 years. Indeed, after Jordan's memoir, several authors gave new proofs of his theorem. The first of these appears to be Blichfeldt, who gave an entirely different proof of Jordan's result via the study of the p-Sylow subgroups, for which he established explicit bounds on their size in terms of p and n (see [4; 5; 6; 7; 35]). Subsequently Bieberbach [3] came up with yet another very different and purely geometric argument, which was later refined by Frobenius [23]. This third proof is much slicker and it is the one that can be found, with some variants, in most textbooks that treat the question, such as [19, Chapter V; 41, Chapter 8]. Blichfeldt himself later combined it with his previous approach to improve his bounds on $J(n)$; see [35] and [46; 26; 22].

Bieberbach's argument starts with what people refer to nowadays as Weyl's unitary trick (i.e., the observation that a finite, or compact, subgroup G of $\mathrm{GL}_n(\mathbb{C})$ can be conjugated inside the compact unitary group $\mathbb{U}_n(\mathbb{C})$ by averaging a hermitian product over G). Then one makes use of a volume packing argument in combination with the *commutator shrinking property* of Lie groups, i.e., the fact that commutators of elements close to the identity in $\mathbb{U}_n(\mathbb{C})$ are themselves close, and in fact much closer, to the identity. This commutator shrinking property has inspired several other authors [50; 10; 31; 1] and is nowadays a crucial tool in the study of discrete subgroups of Lie groups and in Riemannian geometry. We refer the reader to [42, Theorem A] or [14, §2] for a proof of Jordan's theorem via this argument.

Jordan's original proof, on the other hand, was based on a purely algebraic idea, which should be traced back to Klein's method for the classification of the finite subgroups of rotations of the 2-sphere (and isometries of Plato's solids), as described in Klein's famous book on the icosahedron [32]. Basically, one enumerates the elements of G according to the shape and size of their centralizers and one can thus write a class equation involving the order of G and of the centralizers of its elements. Inducting on dimension, this yields a diophantine equation of the form

$$\frac{1}{g} = \frac{1}{q_1} + \cdots + \frac{1}{q_k} - \frac{b}{a}, \tag{1-1}$$

where $g = [G : \Phi]$ is the index of the center Φ of G in G, a, b and k are integers that are bounded in terms of n only and each q_i is the cardinality of a certain subgroup of G/Φ. It is easy to check that any equation of this form forces g to be bounded (in terms of n) and Jordan then discusses the boundedly many cases that may arise. Although more cumbersome, this method gives potentially much more information on the finite subgroup G. For example, Jordan used it to list all finite subgroups of $\mathrm{GL}_3(\mathbb{C})$, giving an explicit set of generators for each class of groups, after examining some 47 different cases.[2]

[2]In fact Jordan missed some groups; see [4; 5; 6; 7; 20].

With the advent of the classification of finite simple groups, B. Weisfeiler [49] and more recently M. Collins [16] have found tight bounds for $J(n)$. For example, Collins proved that if $n \geqslant 71$, then $J(n)$ can be taken to be $(n+1)!$. This is tight, i.e., $(n+1)!$ is always a lower bound for $J(n)$, because the symmetric group on $n+1$ letters acts irreducibly on the hyperplane $\sum_{i=1}^{n+1} x_i = 0$ by permuting the $n+1$ coordinates.

Schur [44] extended Jordan's theorem, proving that it holds assuming only that the group G is torsion (i.e., every element has finite order). In particular, every finitely generated torsion subgroup of $\mathrm{GL}_n(\mathbb{C})$ is finite. This is sometimes called the Jordan–Schur theorem; see [47; 19].

In another direction initiated by Brauer and Feit [13], Larsen and Pink [34] gave a vast generalization of Jordan's theorem to finite linear groups in characteristic p, which avoids the classification of finite simple groups. Interestingly enough, part of their proof is very much akin to Jordan's original argument. See also [11] for a recent use of theorems of Jordan and Larsen–Pink type in the study of finite group actions on elementary abelian p-groups with finite Morley rank.

Finally, we mention that there are nonlinear analogues of Jordan's theorem for finite subgroups of homeomorphisms of manifolds (conjectured by E. Ghys) and for finite subgroups of birational automorphisms of algebraic varieties. In these cases, it has recently been shown that there is a nilpotent subgroup of index bounded only in terms of the dimension of the manifold or variety. But "nilpotent" cannot be replaced by "abelian". We refer the reader to the preprints [18; 25] and references therein for these exciting recent developments.

This paper is organized as follows: in Section 2 we define the notion of M-fan and state the version of Theorem 1.1 that will be used as induction hypothesis. In Sections 3 and 4 we write the corresponding class equation and complete Jordan's proof of Theorem 1.1. In Section 5 we give an illustration of the method by specializing to the case of $n = 2$ and we derive the classical results of Klein for finite subgroups of $\mathrm{SO}_3(\mathbb{R})$. In Section 6 we discuss a nonstandard treatment of Jordan's proof, which is very close to Jordan's original formulation of his proof, and in the last section we briefly survey bounds for $J(n)$ from a historical perspective.

2. A reformulation of Theorem 1.1

As we will see below, Jordan's argument uses nothing about the field $\mathbb{C}$ and in fact his proof carries over to an arbitrary field provided we assume that every element of G is semisimple, i.e., diagonalizable in some field extension. So we let K be an arbitrary field, which we assume algebraically closed without loss of generality.

Let us first reformulate Theorem 1.1 in the form originally proved by Jordan. For this we need to introduce a couple of definitions.

Definition 2.1. By a *root torus*, we mean a subgroup of $\mathrm{GL}_n(K)$ which is conjugate to a subgroup of the diagonal matrices defined by a set of equalities between the diagonal entries.

For example, the subgroup of diagonal matrices

$$\{g = \mathrm{diag}(a_1, \ldots, a_6) \mid a_i \in K^*, a_1 = a_2, a_5 = a_6\}$$

is a root torus of $\mathrm{GL}_6(K)$.

Definition 2.2. Let G be a finite subgroup of $\mathrm{GL}_n(K)$. Given $M \geqslant 2$, we say that a subgroup F of G is an *M-fan* if it is conjugate to a subgroup of the diagonal matrices $\{g = \mathrm{diag}(a_1, \ldots, a_n) \mid a_i \in K^*\}$ such that for every pair of indices i, j the set of ratios $a_i(g)/a_j(g)$ is either reduced to $\{1\}$ or achieves at least M distinct values as g varies in F.

The terminology *fan* is a liberal translation of Jordan's *faisceau*.[3] Note that the subgroup Φ of all scalar matrices in G is clearly an M-fan, for any $M \geqslant 2$.

Note that every M-fan F is contained in a unique minimal root torus S_F defined by the same equalities between diagonal elements, such as $a_i = a_j$, as those that hold in F. In particular $G \cap S_F$ is itself an M-fan and every maximal M-fan in G has this form.

We can now state an alternative, slightly more precise, version of Theorem 1.1:

Theorem 2.3 (Jordan's theorem, second form). *Given $n \in \mathbb{N}$, there are constants $M = M(n)$, $N = N(n) \geqslant 1$ such that the following holds. Let K be an algebraically closed field, and let G be a finite subgroup of $\mathrm{GL}_n(K)$ such that every element of G is diagonalizable. Then G contains a unique maximal M-fan. Call it $\mathcal{F}$. We have $[G : \mathcal{F}] \leqslant N$.*

The proof of Theorem 2.3 spans the next two sections. Before we start, a number of simple remarks are in order:

(1) Since $\mathcal{F}$ is unique, it must be normal in G.

(2) To see that Theorem 2.3 implies Theorem 1.1, it only remains to check that if $K = \mathbb{C}$, then every element of G is diagonalizable. This is indeed the case, because every element of G has finite order and is thus diagonalizable over $\mathbb{C}$.

(3) Although we prove the result in any characteristic, it is worth mentioning that the case of positive characteristic follows from the case when $K = \mathbb{C}$, because if G is as in Theorem 2.3, then $|G|$ is prime to $\mathrm{char}(K)$ and thus G admits an embedding in $\mathrm{GL}_d(\mathbb{C})$. See for instance [39, Proof of Theorem C] or [21, Theorem 3.8].

[3] We are grateful to the referee for suggesting this translation. In fact the word *faisceau* is used throughout Jordan's other works to mean sometimes "subgroup", sometimes "abelian subgroup".

(4) The proof of Theorem 2.3 proceeds by induction on the dimension. The letter F denotes a fan and we reserve the letter $\mathcal{F}$ for maximal fans.

(5) Since $S_{\mathcal{F}}$ is normalized by G, G must permute the eigenspaces of $S_{\mathcal{F}}$. So if G acts primitively on K^n (i.e., does not permute the components of any nontrivial direct sum decomposition of K^n), then $S_{\mathcal{F}}$ must be reduced to scalar matrices and those have bounded index in G.

(6) If $g \in \mathrm{GL}_n(K)$ normalizes $\mathcal{F}$, then it must normalize the root torus $S_{\mathcal{F}}$ too. In particular G lies in the normalizer of a root torus $S_{\mathcal{F}}$ and $[G : G \cap S_{\mathcal{F}}] \leqslant N$.

(7) The abelian normal subgroup A in Theorem 1.1 can be taken to be *characteristic* in G. A theorem of Chermak–Delgado [27, Theorem 1.41] asserts that in any finite group G with an abelian subgroup of index i, there is an abelian characteristic subgroup of index at most i^2. So we could make A characteristic at the expense of changing $J(n)$ into $J(n)^2$. Another route is to observe that, if $M > n!$, the maximal M-fan $\mathcal{F}$ in Theorem 2.3 commutes with every normal abelian subgroup of G. Hence the subgroup generated by all $\alpha(\mathcal{F})$, $\alpha \in \mathrm{Aut}(G)$, is abelian, characteristic, and of smaller index.

As seen from items (1) and (2) above, Theorem 2.3 implies Theorem 1.1. It turns out that one can also derive Theorem 2.3 from Theorem 1.1 directly and we explain this in the paragraph below. To be more precise, since we have only stated Theorem 1.1 over $\mathbb{C}$ while Theorem 2.3 is also valid in positive characteristic, we are going to prove that Theorem 2.3 follows from the assertion that any finite subgroup of $\mathrm{GL}_n(K)$ made of diagonalizable elements admits a normal abelian subgroup of index at most $J(n)$. Jordan's original proof goes by proving Theorem 2.3 first, because its formulation is more adequate for the induction scheme.

Proof of the equivalence of Theorems 1.1 and 2.3. Assume the conclusion of Theorem 1.1. Since every element of G is diagonalizable and A is abelian, A is simultaneously diagonalizable and K^n decomposes as a direct sum of weight spaces (i.e., joint eigenspaces) for A. Since A is normal in G, these eigenspaces are permuted by G and thus G lies in the normalizer $N(S)$ of the root torus S that acts on K^n by a scalar multiple on each one of the weight spaces of A. Note that $A \leqslant S$. Moreover, if F is an M-fan with $M > J(n)$ and $m := [F : F \cap S] \leqslant [G : G \cap S] \leqslant J(n)$, we have $f^m \in S$ for all $f \in F$. Thus $F \cap S$ is an M/m-fan lying in S. Since $M/m > 1$, this implies that F itself lies in S. Hence every M-fan is contained in S. Finally, viewing S as a diagonal subgroup it is straightforward to check that the subgroup generated by all M-fans in S is itself an M-fan. Hence it is the unique maximal M-fan in G. But $G \cap S$ contains some M-fan with index at most $(M-1)^{n-1}$ as follows by intersecting the kernels of the homomorphisms $g \mapsto a_i(g)/a_j(g)$ at most $n-1$ times. Hence $[G : \mathcal{F}] \leqslant J(n)(M-1)^{n-1}$. This completes the claims of Theorem 2.3 with $N = J(n)^n$, $M = J(n) + 1$. $\square$

3. Jordan's fundamental equation

In this section we begin the proof of Theorem 2.3 and obtain Jordan's fundamental equation (3-3) below, which expresses an enumeration of the elements of G into various classes, which we are about to describe. The proof of Theorem 2.3 will be completed in the next section after a discussion of the fundamental equation.

We proceed by induction on the dimension n.

If $n = 1$, then $\mathrm{GL}_1(K) = K^*$ is abelian and there is nothing to prove. We now assume the theorem proven for all dimensions $< n$.

Observe that, by the argument at the end of the last section, it is enough to establish the conclusion of Theorem 1.1, namely the existence of an abelian normal subgroup of index bounded by some function $J(n)$, as this automatically implies the conclusion of Theorem 2.3 with $N(n) = J(n)^n$ and $M(n) = J(n) + 1$.

If G preserves a direct sum decomposition $K^n = K^r \oplus K^{n-r}$, with $1 < r < n$, then we may use the induction hypothesis in the obvious way applying it to the projections $\pi_r(G)$ and $\pi_{n-r}(G)$ to $\mathrm{GL}_r(K)$ and $\mathrm{GL}_{n-r}(K)$, respectively. The conclusion of Theorem 1.1 then easily follows as soon as $J(n) \geqslant \max_{0<r<n} N(r)N(n-r)$ and that of Theorem 2.3 too as we have just said.

We repeatedly use the last observation for subgroups of G that preserve such a decomposition. If $g \in G$ is not a scalar matrix, then the centralizer $C_G(g)$ preserves the eigenspace decomposition of g on K^n. We can therefore apply this observation to $C_G(g)$ and conclude from the induction hypothesis that $C_G(g)$ contains a unique maximal M-fan (for all M larger than a number depending on n only). That is:

Lemma 3.1. *If $g \in G$ is not a scalar matrix, then the centralizer $C_G(g)$ contains a unique maximal M-fan.*

We can thus set the following definition:

Definition 3.2. An element g is said to be *associated with* an M-fan F if F lies in the centralizer $C_G(g)$ and is the unique maximal M-fan of $C_G(g)$.

We denote by F_g the M-fan associated with g. This definition makes sense (so far, thanks to the induction hypothesis) as soon as g is not a scalar matrix in $\mathrm{GL}_n(K)$ by the remarks above the definition. Note that, by maximality, F_g must contain the subgroup Φ of G of all scalar matrices in G. Moreover, setting $N := N(n-1)^2$, it follows from the induction hypothesis that

$$[C_G(g) : F_g] \leqslant N. \tag{3-1}$$

These remarks also have the following three consequences:

Lemma 3.3. *If F is an M-fan of G not entirely made of scalar matrices, then F is contained in a unique maximal M-fan $\mathcal{F}$ of G.*

Proof. Let $f \in F$ be a nonscalar element. If F_1 is an M-fan containing F, then F_1 must commute with all elements of F and thus lie in $C_G(f)$, the centralizer of f. Therefore F_1 must lie in the unique maximal M-fan of $C_G(f)$. □

Let F be an M-fan of G not contained in the scalar matrices Φ and let $\mathcal{F}$ be the maximal M-fan of G containing F. Since $\mathcal{F}$ is contained in the centralizer $C_G(F)$, it must be the maximal M-fan there too and, by the induction hypothesis, we must have $[C_G(F) : \mathcal{F}] \leqslant N$.

Lemma 3.4. *Suppose $\Phi \subsetneqq F \subsetneqq \mathcal{F}$. Then the number n_F of elements of G associated with F is divisible by $|\mathcal{F}|$ and $n_F/|\mathcal{F}| \leqslant N$.*

Proof. If $n_F = 0$ there is nothing to prove, so we assume $n_F \geqslant 1$. Every element associated with F lies in the centralizer $C_G(F)$. Moreover, if $g \in C_G(F)$ is associated with F and $f \in \mathcal{F}$, then gf is also associated with F, i.e., $F_{gf} = F_g = F$. Indeed, since $F \subsetneqq \mathcal{F}$ we must have $gf \notin \Phi$ (as otherwise $C_G(g) = C_G(f)$ contains $\mathcal{F}$) and by Lemma 3.1 there is a unique maximal M-fan F_{gf} in $C_G(gf)$. Since $F \subset C_G(gf)$ we have $F \subset F_{gf} \subset C_G(gf)$. Moreover, F_{gf} is contained in $\mathcal{F}$ and must therefore commute with f, and hence also with g. It follows that $F \subset F_{gf} \subset C_G(g)$ and $F = F_{gf}$ by maximality of F.

Consequently, the set of elements of G associated with F is a union of cosets of $\mathcal{F}$ all lying in $C_G(F)$. Since $C_G(F)$ contains $\mathcal{F}$ as a subgroup of index at most N the result follows. □

And for maximal fans we have:

Lemma 3.5. *Let $\mathcal{F} \neq \Phi$ be a maximal M-fan in G. Then the number $n_{\mathcal{F}}$ of nonscalar elements g in G which are associated with $\mathcal{F}$ is $n_{\mathcal{F}} = |C_G(\mathcal{F})| - |\Phi|$, and $[C_G(\mathcal{F}) : \mathcal{F}] \leqslant N$.*

Proof. A nonscalar element g is associated with $\mathcal{F}$ if and only if $\mathcal{F} \leqslant C_G(g)$, i.e., $g \in C_G(\mathcal{F})$. The bound follows from (3-1). □

The strategy of Jordan's proof consists in enumerating the elements of G according to their associated M-fan. Let Φ be the scalar matrices in G. We may decompose G as the *disjoint* union

$$G = \Phi \cup_F \{g \mid g \text{ associated with } F\},$$

where the union is taken over fans arising as maximal M-fans of centralizers of nonscalar elements of G. We split this union into four disjoint parts,

$$G = \Phi \cup G_1 \cup G_2 \cup G_3,$$

where G_1 is the subset of those g's not in Φ such that $F_g = \Phi$, and G_2 is the subset of those g's not in Φ such that F_g contains Φ strictly but is not the maximal M-fan $\mathcal{F}_g$ which contains it by Lemma 3.3, and finally G_3 is the remaining subset of those

g's not in Φ for which F_g is not Φ and is maximal in G. We now consider each subset G_i one after the other.

(1) We first enumerate the elements of G_1, that is, the g's outside Φ which are associated with Φ. This subset is invariant under conjugation by G. Also it is clearly a union of cosets of Φ, for if $\phi \in \Phi$, then $C_G(g\phi) = C_G(g)$ and thus $g\phi$ is also associated with Φ. It follows that conjugation by G permutes those Φ-cosets.

The stabilizer $N_G(g\Phi)$ of a Φ-coset $g\Phi$ under the G-action by conjugation must contain $C_G(g)$ as a subgroup of index at most n. Indeed, if $h \in G$ has $hg\Phi h^{-1} = g\Phi$, then $hgh^{-1} = g\phi$ for some $\phi \in \Phi$. It follows that $\det(\phi) = 1$ and thus ϕ is an n-th root of unity. We conclude that $[N_G(g\Phi) : C_G(g)] \leqslant n$.

Thus the number of elements in the G-conjugacy class of the coset $g\Phi$ equals

$$\frac{|G|}{|N_G(g\Phi)|}|\Phi| = |G|\frac{1}{[N_G(g\Phi) : C_G(g)]}\frac{1}{[C_G(g) : \Phi]} = |G|\frac{1}{\lambda}.$$

Enumerating all such conjugacy classes, we find

$$|G_1| = |G|\Big(\frac{1}{\lambda_1} + \cdots + \frac{1}{\lambda_{k_1}}\Big),$$

where each λ_i is a positive integer of size at most nN by (3-1) and the remark above.

(2) We now pass to the subset G_2. Clearly G_2 is stable under conjugation by G. Let F be an M-fan of G with maximal M-fan $\mathcal{F}$ such that $\Phi \subsetneqq F \subsetneqq \mathcal{F}$. Let n_F be the number of g's which are associated with F. By Lemma 3.4, $n_F/|\mathcal{F}|$ is an integer of size at most N.

Grouping together the fans that are conjugate to F, we obtain $|G|/|N_G(F)|$ different fans, where $N_G(F)$ is the normalizer of F in G. Note that

$$[N_G(F) : C_G(F)] \leqslant n! \tag{3-2}$$

since $N_G(F)$ permutes the weight spaces of F and hence a subgroup of index at most $n!$ will preserve them and thus commute with F.

It follows that the number of elements that are associated with a fan lying in the G-conjugacy class of F equals

$$n_F\frac{|G|}{|N_G(F)|} = |G|\frac{1}{[N_G(F) : C_G(F)]}\frac{n_F/|\mathcal{F}|}{[C_G(F) : \mathcal{F}]} = |G|\frac{\nu}{\mu},$$

and thus enumerating the different conjugacy classes

$$|G_2| = |G|\Big(\frac{\nu_1}{\mu_1} + \cdots + \frac{\nu_{k_2}}{\mu_{k_2}}\Big),$$

where the $\nu_i \leqslant N$ and $\mu_i \leqslant n!N$ are positive integers.

(3) Finally we consider the subset G_3 of those nonscalar g's such that F_g is maximal in G and different from Φ. Clearly this set is invariant under conjugation by G. Given a maximal M-fan $\mathcal{F}$, the number $n_{\mathcal{F}}$ of elements of G which are associated with $\mathcal{F}$ equals $|C_G(\mathcal{F})| - |\Phi|$ according to Lemma 3.5.

Setting $\omega = [N_G(\mathcal{F}) : C_G(\mathcal{F})]$ and $q = [N_G(\mathcal{F}) : \Phi]$, the number of elements that are associated with a maximal fan conjugate to $\mathcal{F}$ is

$$n_{\mathcal{F}} \frac{|G|}{|N_G(\mathcal{F})|} = |G| \Big(\frac{1}{\omega} - \frac{1}{q} \Big),$$

where ω and q are positive integers with $\omega \leqslant n!$ and $q = [C_G(\mathcal{F}) : \Phi]\omega \geqslant 2\omega$.

Summing over the conjugacy classes, we get

$$|G_3| = |G| \Big(\Big(\frac{1}{\omega_1} - \frac{1}{q_1} \Big) + \cdots + \Big(\frac{1}{\omega_{k_3}} - \frac{1}{q_{k_3}} \Big) \Big).$$

Combining all three cases, we have thus completed our enumeration of G and we obtain:

Proposition 3.6 (Jordan's fundamental equation). *Let G be a finite subgroup of $\mathrm{GL}_n(K)$ all of whose elements are diagonalizable, and Φ the subgroup of scalar matrices in G. Then there are positive integers q_i dividing $g := |G|/|\Phi|$ such that*

$$|G| = |\Phi| + |G| \sum_{i=1}^{k_1} \frac{1}{\lambda_i} + |G| \sum_{i=1}^{k_2} \frac{\nu_i}{\mu_i} + |G| \sum_{i=1}^{k_3} \Big(\frac{1}{\omega_i} - \frac{1}{q_i} \Big), \tag{3-3}$$

where k_i, λ_i, ν_i, μ_i and ω_i are nonnegative integers of size at most $2n!N$ (recall that $N = N(n-1)^2$ is the bound from Theorem 2.3 under the induction hypothesis). In particular,

$$\frac{1}{g} = \frac{1}{q_1} + \cdots + \frac{1}{q_{k_3}} - \frac{b}{a}, \tag{3-4}$$

where $\frac{b}{a}$ is an irreducible fraction whose numerator and denominator are bounded in terms of n only.

To prove Proposition 3.6 it remains only to show the bound on the number k_i of elements in each sum and then derive (3-4). But this follows from the equation (3-3) and from the bounds previously obtained, because

$$\frac{1}{\omega_i} - \frac{1}{q_i} \geqslant \frac{1}{2\omega_i}$$

for each $i = 1, \ldots, k_3$ and thus each term in the above sums contributes at least $|G|/2n!N$, forcing $k_1 + k_2 + k_3 \leqslant 2n!N$.

Showing (3-4) is a simple matter of rearranging (3-3):

$$\frac{|G|}{|\Phi|}\left(\sum_{i=1}^{k_1}\frac{1}{\lambda_i}+\sum_{i=1}^{k_2}\frac{\nu_i}{\mu_i}+\sum_{i=1}^{k_3}\frac{1}{\omega_i}-1\right)=\sum_{i=1}^{k_3}\frac{|G|}{|\Phi|}\frac{1}{q_i}-1. \tag{3-5}$$

Then we let

$$g:=\frac{|G|}{|\Phi|} \quad \text{and} \quad \frac{b}{a}:=\sum_{i=1}^{k_1}\frac{1}{\lambda_i}+\sum_{i=1}^{k_2}\frac{\nu_i}{\mu_i}+\sum_{i=1}^{k_3}\frac{1}{\omega_i}-1,$$

where $\frac{b}{a}$ is an irreducible fraction. We thus get (3-4).

Note further that a is bounded in terms of n only: indeed it cannot exceed the least common multiple of at most $2n!N$ integers of size at most $n!N$. A similar bound holds for b. This completes the proof of Proposition 3.6.

4. Proof of Theorem 2.3

It remains to discuss the fundamental equation (3-3) according to the possible values of the integers λ_i, ν_i, μ_i, ω_i and q_i.

The proof rests on the following elementary lemma about fractions:

Lemma 4.1. *Consider the following equation, where all variables are positive integers*:

$$\frac{1}{g}=\frac{1}{q_1}+\cdots+\frac{1}{q_k}-\frac{b}{a}.$$

Suppose that $q_i<g$ for all i. Then $g\leqslant f(k,a)$, where $f(k,a)$ is a function of k and a only. One may take $f(k,a)=(k!a)^{2^k}$.

Proof. The proof proceeds by induction on k. If $k=1$, then $\frac{1}{q_1}\geqslant\frac{b}{a}$ implies $q_1\leqslant a$ and $\frac{1}{g}\geqslant\frac{1}{q_1 a}\geqslant\frac{1}{a^2}$, so $g\leqslant a^2=:f(1,a)$.

Suppose the lemma proven for all indices $\leqslant k-1$. Without loss of generality, we may assume that $\frac{1}{q_1}\leqslant\ldots\leqslant\frac{1}{q_k}$. Then $\frac{1}{q_2}+\cdots+\frac{1}{q_k}<\frac{b}{a}$, for $q_1<g$.

It follows that

$$\frac{b}{a}>\frac{1}{q_2}+\cdots+\frac{1}{q_k}\geqslant\frac{1}{q_k}.$$

We may thus write $\frac{c}{d}=\frac{b}{a}-\frac{1}{q_k}$, where c,d are positive integers and $\frac{c}{d}$ is an irreducible fraction.

Since $\frac{1}{g}\leqslant\frac{k}{q_k}-\frac{b}{a}$, we get $q_k\leqslant ka$ and thus $d=\operatorname{lcm}(a,q_k)\leqslant ka^2$. We obtain

$$\frac{1}{g}=\frac{1}{q_1}+\cdots+\frac{1}{q_{k-1}}-\frac{c}{d}.$$

Applying the induction hypothesis we conclude $g\leqslant f(k-1,ka^2)=:f(k,a)$. $\square$

We now complete the proof of Theorem 2.3. If $k_3 = 0$, then we see from (3-4) that $\frac{-b}{a} = \frac{1}{g}$ so $b = -1$ and $g = a$ is bounded in terms of n only by Proposition 3.6. Hence Φ has bounded index in G and we are done.

Assume $k_3 \geqslant 1$. If $q_i = g$ for some i, then

$$G = N_G(\mathcal{F}_i) \quad \text{and} \quad [G : \mathcal{F}_i] = \omega_i [C_G(\mathcal{F}_i), \mathcal{F}_i] \leqslant n!N$$

by (3-1). So $\mathcal{F}_i$ is the desired abelian normal subgroup of bounded index and we are done.

The right-hand side of (3-5) is a nonnegative positive integer. If it is zero, then $k_3 = 1$, $q_1 = g$ and we fall back to the previous case. Otherwise it is positive and thus $b > 0$, so that we are in the situation of Lemma 4.1. We conclude that g is bounded in terms of n only and again we are done.

Theorem 2.3 is now proven in full.

Remark. We mention in passing that the proof of Landau's theorem [33] that there are only finitely many finite groups G with exactly k conjugacy classes $\mathrm{cl}_1, \dots, \mathrm{cl}_k$ is based on a similar, and easier, diophantine equation, namely

$$1 = \frac{1}{q_1} + \cdots + \frac{1}{q_k},$$

where $q_i = |G|/|\mathrm{cl}_i|$. This is an instance of an *Egyptian fraction* [9], and a simple argument [36] implies that

$$k \geqslant \frac{1}{\log 4} \log \log |G|.$$

5. Platonic solids and the finite subgroups of $\mathrm{SO}_3(\mathbb{R})$

As an illustration and for the sake of comparison, we recall in this section a proof of the classification of finite subgroups of $\mathrm{SO}_3(\mathbb{R})$ following Klein's method, as given in many textbooks, e.g., [45; 51].

Let G be a finite subgroup of $\mathrm{SO}_3(\mathbb{R})$. Every nontrivial element of G is a rotation around some axis. Let X be the set of all axes that arise as axes of rotations in G. Clearly G permutes X because if x_h is the axis of $h \in G$, then $gx_h = x_{ghg^{-1}}$. The determination of all possible groups G proceeds via a double counting argument, or class equation, which enumerates the elements of G according to their fixed axis. Given an axis $x \in X$, let G_x be the subset of elements of G whose axis is x, to which we adjoin the identity. Then G_x is a subgroup. It usually coincides with the centralizer of x, except when x is a flip (i.e., has angle π). Enumerating the elements of G starting with the identity element, we can write

$$|G| = 1 + \sum_{x \in X} (|G_x| - 1).$$

To go further, we group together the terms corresponding to two axes that are G-congruent (i.e., $x \sim y$ if there is $g \in G$ with $y = gx$). We obtain

$$|G| = 1 + \sum_{\text{classes of } x \in X} \frac{|G|}{|\mathrm{Stab}_x|}(|G_x| - 1), \tag{5-1}$$

where Stab_x is the stabilizer of x in G. It is a subgroup of G. Now observe that an element g of G which preserves x may be only of two possible forms: either it fixes both poles of x, in which case g belongs to G_x, or it permutes the two poles of x. It follows that G_x is a subgroup of Stab_x of index either 1 or 2. Let $x_1, \dots, x_r, x_{r+1}, \dots, x_{r+s}$ be a set of representatives of the G-orbits in X such that $[\mathrm{Stab}_{x_i} : G_{x_i}] = 1$ if $1 \leqslant i \leqslant r$ and $[\mathrm{Stab}_{x_i} : G_{x_i}] = 2$ if $r+1 \leqslant i \leqslant r+s$. Setting $g_i = |G_{x_i}|$, dividing by $|G|$ in (5-1) we obtain

$$1 = \frac{1}{|G|} + \sum_{i=1}^{r}\Big(1 - \frac{1}{g_i}\Big) + \frac{1}{2}\sum_{i=r+1}^{r+s}\Big(1 - \frac{1}{g_i}\Big), \tag{5-2}$$

or equivalently, in the form of Jordan's fundamental equation (3-4),

$$\frac{1}{|G|} = \frac{1}{n_1} + \cdots + \frac{1}{n_{r+s}} - \frac{b}{a}, \tag{5-3}$$

where $\frac{b}{a} = r + \frac{s}{2} - 1$, and $n_i = g_i$ for $i \leqslant r$, $n_i = 2g_i$ for $i > r$.

It remains to discuss equation (5-3) according to the possible values of the g_i's. Since $g_i \geqslant 2$, we get from (5-2) that $1 > \frac{r}{2} + \frac{s}{4}$, from which it follows immediately that $r \leqslant 1$ and $2r + s \leqslant 3$, so $\frac{b}{a} \in \{-\frac{1}{2}, 0, \frac{1}{2}\}$. Since n_i divides $|G|$, (5-3) forces $\frac{b}{a} > 0$ (hence $\frac{b}{a} = \frac{1}{2}$), unless $\frac{b}{a} = 0$ and $r + s = 1$. This last case can only occur if $r + \frac{s}{2} = 1$, forcing $r = 1, s = 0$. We now examine the various possibilities.

- $\frac{b}{a} = 0$ and $r = 1$, $s = 0$. Then $g_1 = |G|$ and G is a *cyclic group* of rotations around a single axis.
- $\frac{b}{a} = \frac{1}{2}$, and $r = 1$, $s = 1$,

$$\frac{1}{2} + \frac{1}{|G|} = \frac{1}{n_1} + \frac{1}{n_2}.$$

Since $g_2 \geqslant 2$, we have $n_2 \geqslant 4$. This forces $n_1 < 4$ and hence $n_1 = 2, 3$. There are thus two cases:

(1) If $n_1 = 2$, then $n_2 = 2g_2 = |G|$ and G is a *dihedral group* of order $2n$, with $n = g_2$ an *odd integer*. G is the group of orientation preserving isometries of a regular n-gon. Moreover in our case n is *odd* because there are only two G-orbits of axes.

(2) If $n_1 = 3$, then one checks that $n_2 = 4$ and $|G| = 12$. Here G is the group of orientation preserving isometries of a *regular tetrahedron*.

- $\frac{b}{a} = \frac{1}{2}$, $r = 0$, $s = 3$,

$$\frac{1}{2} + \frac{1}{|G|} = \frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{n_3}.$$

We may assume $n_1 \leqslant n_2 \leqslant n_3$. This forces $n_1 < 6$. But $n_1 = 2g_1 \geqslant 4$. So $n_1 = 4$ and thus

$$\frac{1}{4} + \frac{1}{|G|} = \frac{1}{2g_2} + \frac{1}{2g_3}.$$

We have the following cases:

(1) $g_2 = 2$, then $|G| = n_3 = 2g_3$ and G is a *dihedral group* of order $2n$, with $n = g_3$ an *even integer*. G is the group of orientation preserving isometries of a regular n-gon. Moreover in our case n is *even* because there are exactly three G-orbits of axes. For example if $n = 2$, $G \simeq (\mathbb{Z}/2\mathbb{Z})^2$ and every nontrivial element is a flip around one of three mutually orthogonal axes.

(2) $g_2 = 3$, then

$$\frac{1}{6} + \frac{2}{|G|} = \frac{1}{g_3}.$$

This forces $g_3 < 6$, and hence three cases:

(a) $g_3 = 3$, then $|G| = 12$ and G preserves a regular tetrahedron. This however is in contradiction with the assumption $s = 3$, since there are only two G-orbits of axes in this case. So this case cannot occur.

(b) $g_3 = 4$, then $|G| = 24$ and $G \simeq S_4$ is the group of orientation preserving isometries of a *cube* or regular *octahedron*.

(c) $g_3 = 5$, $G \simeq A_5$ is the group of orientation preserving isometries of a regular *icosahedron* or *dodecahedron*.

6. Nonstandard analysis and Jordan's unlimited numbers

As we have seen, Jordan gave no explicit bound on $J(n)$ in his article. Of course, this is not due to any fundamental ineffectiveness in the proof. Indeed, if one very carefully follows Jordan's argument, then it is possible to obtain in this way a bound in the form of a tower of exponentials, i.e., a tower

$$10^{10^{\cdot^{\cdot^{\cdot^{10}}}}}$$

of length n; see [2]. In fact Jordan himself seems to have been dissatisfied with his original exposition and devoted a second article [30], where he rewrote his proof and explained why it is effective (even though he still did not supply a concrete bound). From a purely epistemological viewpoint, it is however interesting to consider how Jordan gets away with not writing down any bound whatsoever in his original memoir. In fact, in order to convince the reader of the soundness of his argument, he introduces a distinction between two kinds of numbers, which he calls limited and unlimited. Let us quote him [29, p. 114]:

> It is important for the study thereafter to make precise the meaning we attach to the words *limited* and *unlimited*. They are not synonymous to *finite* and *infinite*. We will say that a number is limited if it is smaller than a certain bound that has been determined. It follows from this definition that a finite number, about which we have no data, is unlimited; but it becomes limited as soon as we manage to assign a bound to it.

This way, instead of saying that a certain quantity is bounded in terms of n only, he says that the quantity is limited, while if it is not, it is unlimited and this is somehow leading to a contradiction when considering the class equation (3-4). A century and a half after Jordan, it is hard not to see there the premise of a way of thinking that prefigures nonstandard analysis, where a new kind of number, the unlimited ones, is given an existence of its own.

In fact, it is possible to give a nonstandard treatment of Jordan's proof, which we now sketch. Starting with a sequence of possible counterexamples to the theorem, one may take their ultraproduct, which becomes a certain infinite, pseudofinite subgroup $\widehat{G}$ of $\mathrm{GL}_n(\widehat{K})$. Here $\widehat{K}$ is the ultraproduct of algebraically closed fields K_i. A fan in $\widehat{G}$ is defined to be an internal diagonalizable subgroup F such that each root $\alpha_{ij} : F \to \widehat{K}$, $f \mapsto \lambda_i(f)/\lambda_j(f)$, is either infinite or trivial. The induction hypothesis on the dimension allows one to assume that the centralizer $C(g)$ of every nonscalar element $g \in \widehat{G}$ admits a unique normal maximal fan F_g whose index is finite. Arguing as in Jordan's proof, we can partition the elements of $\widehat{G}$ into four parts: scalars Φ, elements $g \notin \Phi$ with $F_g = \Phi$, elements $g \notin \Phi$ with $\Phi \lneq F_g$ and F_g not maximal in $\widehat{G}$, elements $g \notin \Phi$ with F_g maximal. Exploiting the fact that $\widehat{G}$ is pseudofinite, the second and third parts form a proportion of $r|\widehat{G}|$ of $\widehat{G}$, while the last part forms a proportion $\big(r' - \sum q_i^{-1}\big)|\widehat{G}|$, where r, r' are finite (standard) rationals and $q_i = |N_{\widehat{G}}(F_{g_i})/\Phi|$ are a finite number of (nonstandard) divisors of $|\widehat{G}/\Phi|$. Denoting by g the nonstandard integer $|\widehat{G}/\Phi|$, we thus have the equation

$$\frac{1}{g} = r'' + \sum_i \frac{1}{q_i},$$

where $r'' = 1 - (r + r')$. Taking the difference with standard parts we see that this implies

$$\frac{1}{g} = {\sum_i}' \frac{1}{q_i}$$

for a subsum of the original sum. However, the q_i's are divisors of g and the only way this can happen is if $q_i = g$ for some i. By (3-1) and (3-2), this means that F_{g_i} is normal in $\widehat{G}$ and of finite index. This contradiction ends the proof.

The reader curious to take a look at Jordan's original article will see that the above nonstandard treatment is in fact much closer to Jordan's own formulation of his proof than the exposition we have given of it in Section 3. Indeed Jordan does not talk about M-fans, but only defines fans. And he does so exactly as we did in the nonstandard treatment above only using the word "illimited" in place of the word "infinite". Of course this definition can only make sense rigorously if we place ourselves in a nonstandard universe to begin with. So his proof is resolutely nonstandard since its very inception. His original formulation [29, §40 p. 114] then reads as follows:

Theorem 6.1 (Jordan's theorem, original formulation). *A finite subgroup G of linear substitutions admits a unique maximal fan. It is normal and its index is a limited number.*

To finish, we stress the key role of the finiteness of G in Jordan's theorem. In Jordan's proof it is exploited arithmetically via the class equation. This is to be contrasted with Bieberbach's geometric argument via the commutator shrinking property, where finiteness is exploited via the element closest to the identity.

We end this section by mentioning in passing some related recent developments around a question of Zilber [52, Problem 6.3] regarding pseudofinite groups. Recently, Nikolov, Schneider and Thom [38] proved that every homomorphism from a pseudofinite group to a compact Lie group has abelian-by-finite image, thus answering a conjecture of Pillay [40, Conjecture 1.7] and Zilber's question by the same token.

Of course such a strong statement is more than enough to establish Jordan's theorem itself following the nonstandard approach sketched above, at least when the characteristic of K is zero. Indeed, in this case $\widehat{K}$ can be taken to be isomorphic to $\mathbb{C}$ as any ultraproduct of countable algebraically closed fields of characteristic zero. Furthermore, we may assume that $\widehat{G}$ lies in the internal set of unitary matrices $\mathbb{U}_n(\widehat{K})$, in other words that $\widehat{G}$ is a subgroup of a compact Lie group. By Nikolov–Schneider–Thom, this implies that $\widehat{G}$ is abelian-by-finite, which is the desired contradiction.

Of course the theorem of Nikolov, Schneider and Thom lies much deeper than Jordan's theorem, because it applies to any pseudofinite group and not only those lying in some GL_n for some fixed n. Their proof relies on the deep results of Nikolov and Segal [37] about commutator width in finite groups.

7. Bounds on $J(n)$

To conclude, we briefly survey the history around Jordan's theorem and how bounds on $J(n)$ have sharpened over time:

- Jordan (1878): no bound (in fact: tower of exponentials [2]).
- Blichfeldt (1905): $\exp(O(n^3))$.
- Bieberbach (1911): $(1+32^4n^{10})^{2n^2}$.
- Frobenius (1911): $(\sqrt{8n}+1)^{2n^2}$.
- Blichfeldt (1917): $n!6^{(n-1)(\pi(n+1)+1)}$ ($\pi(x)$ denoting the number of prime numbers $\leqslant x$).
- Weisfeiler (1984): $(n+1)!e^{O(\log n)^2}$ (using CFSG).
- Collins (2007): $(n+1)!$ for $n \geqslant 71$ (using CFSG).

In [4, p. 396] Blichfeldt, who had just completed his dissertation under Sophus Lie, shows that no prime $p \geqslant n(2n-1)$ divides the order g (modulo the center) of a finite primitive subgroup of $\mathrm{GL}_n(\mathbb{C})$ (see also [8, §64]). In [5, p. 321] he obtains bounds for the p-exponent of g. He shows [7, Theorem 16, p. 42] that g is a divisor of $n!(2\cdot 3\cdots\cdot p\cdots\cdot)^{n-1}$, where the product extends over all primes $p < n(2n-1)$. This is of order at most $\exp(O(n^3))$, and this implies a bound of the same magnitude for $J(n)$ for arbitrary finite subgroups (implicit in [5, §12, p. 320] and [6, p. 232]).

In his 1917 monograph [8], Blichfeldt furthers his earlier results, incorporating a geometric argument inspired by Bieberbach's argument [3] and Frobenius' improvement [23]. He shows in [8, §73] that an abelian subgroup of a primitive group must have order at most 6^{n-1} times the size of the group of scalar matrices. This is based on a lemma [8, §70] according to which in a finite primitive subgroup, any transformation whose eigenvalues are concentrated in an arc of length at most $\frac{2\pi}{3}$ centered at one of them on the unit circle, must be scalar. This lemma is closely related to the Bieberbach–Frobenius proofs. And finally he derives the bound $n!6^{(n-1)(\pi(n+1)+1)}$ on $J(n)$, where $\pi(n)$ is the number of primes $\leqslant n$. See [22, Chapter 30] for a thorough treatment of Blichfeldt's bound and [42; 43] for recent improvements on Blichfeldt's lemma. In fact Blichfeldt claimed that 6 could be replaced by 5, but no proof of this has appeared. The three-author book [35], which is dedicated to Camille Jordan, also contains a summary of these results.

Other excellent expositions of Blichfeldt's bound are contained in [46] and [26]. See also [21, Chapter 5] for a treatment of Blichfeldt's earlier results and a proof of Jordan's theorem using the Bieberbach–Frobenius argument.

Brauer [12] conjectured that Blichfeldt's bound could be improved to one of the form $e^{O(n\log n)}$, and indeed he was able to achieve it under certain hypotheses. In fact, for finite solvable subgroups Dornhoff proved an exponential bound $2^{4n/3}3^{10n/9-1/3}$ that is even sharp for infinitely many n's [22, Theorem 36.4].

Nevertheless, Blichfeldt's second bound of the form $e^{O(n^2/\log n)}$ seems to be the best one available without the classification of finite simple groups (CFSG). This small looking gain of a factor $(\log n)^2$ in the exponent compared to the Bieberbach–Frobenius bound can sometimes prove important, as we have found out in [15].

Shortly before disappearing while hiking in Chile,[4] B. Weisfeiler announced a bound on $J(n)$ of $e^{O(n \log n)}$ quality [49]. His unpublished manuscript has now been typed up and is available online [48]. Finally, more recently, M. Collins [16; 17] has improved the bound for n large to one that is sharp, namely $(n+1)!$, thus closing a long chapter in the history of finite linear groups.

Acknowledgements

I am grateful to Udi Hrushovski, who first drew my attention to Jordan's original memoir, to Michael Collins and Terry Tao for useful comments and to Jean-Pierre Serre for encouraging me to publish this note, a first version of which was written some ten years ago. I also thank the referee for their thoughtful comments and careful reading.

References

[1] H. Bass, "Theorems of Jordan and Burnside for algebraic groups", *J. Algebra* **82**:1 (1983), 245–254. MR Zbl

[2] D. Bassett, "Jordan's theorem on finite linear groups", undergraduate research report, University of Cambridge, 2021.

[3] L. Bieberbach, "Über einen Satz des Hrn C. Jordan in der Theorie der endlichen Gruppen linearer Substitutionen", *Sitzber. Preuss. Akad. Wiss.* (1911), 231–240. Zbl

[4] H. F. Blichfeldt, "On the order of linear homogeneous groups", *Trans. Amer. Math. Soc.* **4**:4 (1903), 387–397. MR Zbl

[5] H. F. Blichfeldt, "On the order of linear homogeneous groups, II", *Trans. Amer. Math. Soc.* **5**:3 (1904), 310–325. MR Zbl

[6] H. F. Blichfeldt, "On imprimitive linear homogeneous groups", *Trans. Amer. Math. Soc.* **6**:2 (1905), 230–236. MR Zbl

[7] H. F. Blichfeldt, "On the order of linear homogeneous groups, IV", *Trans. Amer. Math. Soc.* **12**:1 (1911), 39–42. MR Zbl

[8] H. F. Blichfeldt, *Finite collineation groups*, University of Chicago Press, 1917. Zbl

[9] T. F. Bloom and C. Elsholtz, "Egyptian fractions", *Nieuw Arch. Wiskd.* (5) **23**:4 (2022), 237–245. MR

[10] W. M. Boothby and H.-C. Wang, "On the finite subgroups of connected Lie groups", *Comment. Math. Helv.* **39** (1965), 281–294. MR Zbl

[11] A. Borovik, "Finite group actions on abelian groups of finite Morley rank", preprint, 2020. arXiv 2008.00604

[12] R. Brauer, "On the order of finite primitive projective groups in a given dimension", *Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II* **1969** (1969), 103–106. MR Zbl

[4] See https://en.wikipedia.org/wiki/Boris_Weisfeiler.

[13] R. Brauer and W. Feit, "An analogue of Jordan's theorem in characteristic p", *Ann. of Math.* (2) **84** (1966), 119–131. MR Zbl

[14] E. Breuillard and B. Green, "Approximate groups, III: The unitary case", *Turkish J. Math.* **36**:2 (2012), 199–215. MR Zbl

[15] E. Breuillard and G. Pisier, "Random unitaries, amenable linear groups and Jordan's theorem", preprint, 2017. arXiv 1703.07892

[16] M. J. Collins, "On Jordan's theorem for complex linear groups", *J. Group Theory* **10**:4 (2007), 411–423. MR Zbl

[17] M. J. Collins, "Bounds for finite primitive complex linear groups", *J. Algebra* **319**:2 (2008), 759–776. MR Zbl

[18] B. Csikós, L. Pyber, and E. Szabó, "Finite subgroups of the homeomorphism group of a compact topological manifold are almost nilpotent", preprint, 2022. arXiv 2204.13375

[19] C. W. Curtis and I. Reiner, *Representation theory of finite groups and associative algebras*, Pure and Appl. Math. **XI**, Interscience Publishers, New York, 1962. MR Zbl

[20] J. Dieudonné, "Notes sur les travaux de C. Jordan relatifs à la théorie des groupes finis", in *Œuvres de Camille Jordan*, vol. III, Gauthier-Villars & Cie., Paris, 1962. Zbl

[21] J. D. Dixon, *The structure of linear groups*, Van Nostrand Reinhold Math. Studies **37**, Van Nostrand Reinhold Company, London, 1971. Zbl

[22] L. Dornhoff, *Group representation theory*, *Part A*: *Ordinary representation theory*, Pure and Appl. Math. **7**, Marcel Dekker, New York, 1971. MR Zbl

[23] G. Frobenius, "Über den von L. Bieberbach gefundenen Beweis eines Satzes von C. Jordan", *Sitzber. Preuss. Akad. Wiss.* (1911), 241–248. Zbl

[24] J. J. Gray, *Linear differential equations and group theory from Riemann to Poincaré*, 2nd ed., Birkhäuser, Boston, 2000. MR Zbl

[25] A. Guld, "Finite subgroups of the birational automorphism group are 'almost' nilpotent of class at most two", preprint, 2020. arXiv 2004.11715

[26] I. M. Isaacs, *Character theory of finite groups*, Pure and Appl. Math. **69**, Academic Press, New York, 1976. MR Zbl

[27] I. M. Isaacs, *Finite group theory*, Grad. Studies in Math. **92**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[28] M. C. Jordan, "Sur la détermination des groupes formés d'un nombre fini de substitutions linéaires", *Comp. Rend. Acad. Sc.* **83** (1876), 1035–1037.

[29] M. C. Jordan, "Mémoire sur les équations différentielles linéaires à intégrale algébrique", *J. Reine Angew. Math.* **84** (1878), 89–215. MR Zbl

[30] M. C. Jordan, "Sur la détermination des groupes d'ordre fini contenus dans le groupe linéaire", *Atti R. Accad. Sci. Fis. Mat.* **VIII**:11 (1879).

[31] D. A. Každan and G. A. Margulis, "A proof of Selberg's hypothesis", *Mat. Sb.* (*N.S.*) **75 (117)** (1968), 163–168. In Russian; translated in *Math. USSR Sb.* **4** (1969), 147–152. MR Zbl

[32] F. Klein, *Lectures on the icosahedron and the solution of equations of the fifth degree*, revised ed., Dover Publications, New York, 1956. MR Zbl

[33] E. Landau, "Über die Klassenzahl der binären quadratischen Formen von negativer Discriminante", *Math. Ann.* **56**:4 (1903), 671–676. MR

[34] M. J. Larsen and R. Pink, "Finite subgroups of algebraic groups", *J. Amer. Math. Soc.* **24**:4 (2011), 1105–1158. MR Zbl

[35] G. A. Miller, H. F. Blichfeldt, and L. E. Dickson, *Theory and applications of finite groups*, 2nd ed., Stechert, New York, 1938.

[36] M. Newman, "A bound for the number of conjugacy classes in a group", *J. London Math. Soc.* **43** (1968), 108–110. MR Zbl

[37] N. Nikolov and D. Segal, "Generators and commutators in finite groups; abstract quotients of compact groups", *Invent. Math.* **190**:3 (2012), 513–602. MR Zbl

[38] N. Nikolov, J. Schneider, and A. Thom, "Some remarks on finitarily approximable groups", *J. Éc. Polytech. Math.* **5** (2018), 239–258. MR Zbl

[39] M. V. Nori, "On subgroups of $\mathrm{GL}_n(\mathbf{F}_p)$", *Invent. Math.* **88**:2 (1987), 257–275. MR Zbl

[40] A. Pillay, "Remarks on compactifications of pseudofinite groups", *Fund. Math.* **236**:2 (2017), 193–200. MR Zbl

[41] M. S. Raghunathan, *Discrete subgroups of Lie groups*, Ergebnisse der Math. **68**, Springer, 1972. MR Zbl

[42] G. R. Robinson, "On linear groups", *J. Algebra* **131**:2 (1990), 527–534. MR Zbl

[43] G. R. Robinson, "On eigenvalues of elements of finite linear groups", *J. Algebra* **165**:1 (1994), 113–121. MR Zbl

[44] I. Schur, "Über Gruppen periodischer linearer Substitutionen", *Sitzber. Preuss. Akad. Wiss.* (1911), 619–627. Zbl

[45] M. Senechal, "Finding the finite groups of symmetries of the sphere", *Amer. Math. Monthly* **97**:4 (1990), 329–335. MR Zbl

[46] A. Speiser, *Die Theorie der Gruppen von endlicher Ordnung*, Birkhäuser, Basel, 1956. MR Zbl

[47] B. A. F. Wehrfritz, *Infinite linear groups*, Ergebnisse der Math. **76**, Springer, 1973. MR Zbl

[48] B. Weisfeiler, "On the size and structure of finite linear groups", unpublished manuscript, 1984. Zbl arXiv 1203.1960

[49] B. Weisfeiler, "Post-classification version of Jordan's theorem on finite linear groups", *Proc. Nat. Acad. Sci. U.S.A.* **81**:16, , Phys. Sci. (1984), 5278–5279. MR Zbl

[50] H. Zassenhaus, "Beweis eines Satzes über diskrete Gruppen", *Abh. Math. Sem. Univ. Hamburg* **12**:1 (1937), 289–312. MR Zbl

[51] H. Zassenhaus, *The theory of groups*, Chelsea Publishing Co., New York,, 1949. MR Zbl

[52] B. Zilber, "Perfect infinities and finite approximation", pp. 199–223 in *Infinity and truth*, edited by C. Chong et al., Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. **25**, World Sci. Publ., Hackensack, NJ, 2014. MR Zbl

Received 3 Oct 2022. Revised 19 Nov 2022.

EMMANUEL BREUILLARD:

breuillard@maths.ox.ac.uk
Mathematical Institute, University of Oxford, Oxford, United Kingdom

Model Theory
Vol. 2, No. 2, 2023
https://doi.org/10.2140/mt.2023.2.449

Higher internal covers

Moshe Kamensky

We define and study a higher-dimensional version of model-theoretic internality, and relate it to higher-dimensional definable groupoids in the base theory.

1. Introduction

The model-theoretic notions of *internality* and the binding group came up originally in work of Zil'ber on categorical theories [13], and shortly after of Poizat [11] in the ω-stable context, where it was also noticed that differential Galois theory occurs as a special case. The stability hypothesis was completely removed in [3, Appendix B], where it was shown that the crucial hypothesis is *stable embeddedness* of the base theory.

Internality is a condition on a definable set $\boldsymbol{Q}$ in an expansion $\mathcal{T}^*$ of a theory $\mathcal{T}$ to "almost" be interpretable in $\mathcal{T}$: it is interpretable after adding a set of parameters to $\mathcal{T}^*$. In this situation, the theory provides a *definable* group $\boldsymbol{G}$ in $\mathcal{T}^*$, acting definably on $\boldsymbol{Q}$ as its group of automorphisms fixing all elements in the reduct $\mathcal{T}$. It is important here that the binding group $\boldsymbol{G}$ is defined in $\mathcal{T}^*$ rather than in $\mathcal{T}$: in applications, one often understands groups in $\mathcal{T}$ better than in $\mathcal{T}^*$. The group $\boldsymbol{G}$ itself is also internal to $\mathcal{T}$, and as a result can be identified with a definable group $\boldsymbol{H}$ in $\mathcal{T}$, but only noncanonically (and in general, only after adding parameters). In the context of differential Galois theory, this is related to the fact that the group of points of the (algebraic) Galois group of a differential equation does not act on the set of solutions, and its identification with the group of automorphisms is not canonical.

The noncanonicity was explained by Hrushovski in [4], where it is shown that the natural object that appears in this context is a definable *groupoid* in $\mathcal{T}$, with the different groups $\boldsymbol{H}$ occurring as the groups of automorphisms of each object. In fact, it is shown there that there is a correspondence between groupoids definable in the base theory $\mathcal{T}$ and internal sorts in expansions of $\mathcal{T}$. This correspondence is reviewed in Section 2. It is also suggested in [4] that sorts of $\mathcal{T}^*$ internal to $\mathcal{T}$ should be viewed as generalised sorts of $\mathcal{T}$, obtained as a quotient by the

MSC2020: 03C40.
Keywords: internality, definable groupoid, higher groupoid, binding group.

corresponding definable groupoid, just like an imaginary sort is obtained from a definable equivalence relation (which is a special case). In the current paper we try to follow this suggestion, by considering what should be the correct notion of internality, after viewing these new sorts as “legitimate” definable sorts.

Our approach is motivated by topology. There, a typical example of a groupoid arises as the fundamental groupoid $\pi_1(X)$ of a space X, i.e., the groupoid whose objects are the points of X, and whose morphisms are homotopy classes of paths. For sufficiently nice X, this groupoid can be described in terms of the category of *local systems* (locally constant sheaves) on X: each point of X determines a functor to the category of sets, satisfying suitable properties (for example, it commutes with products), and each path determines a map between such functors (which depends only on the homotopy class since the system is locally constant). We propose to view internality as analogous to this picture: definable sets in the theory corresponding to a definable groupoid $\boldsymbol{G}$ in $\mathcal{T}$ can be viewed as local systems (of definable sets) on $\boldsymbol{G}$, and conversely. This point of view is explained in Section 2.2.5 (the base theory $\mathcal{T}$ corresponds to a contractible space in this approach, so definable sets in it correspond to constant systems).

By definition, the local systems on X do not tell us anything about the homotopy type of X above homotopical dimension 1. To encode higher homotopical information, we may try looking at families of spaces rather than of sets. A space X is called *n-truncated* if $\pi_k(X, x)$ is trivial for all $k > n$ and base points $x \in X$. Such spaces are represented in homotopy theory by what we call in this paper n-groupoids (Definition 3.2.3 in the definable setting; these are equivalent to n-categories in the sense of [8, §2.3.4] which are groupoids). In the case $n = 1$, these are usual groupoids, and the previous paragraph suggests studying them by systems of 0-truncated spaces, i.e., sets. Going one dimension higher, one expects to recover 2-groupoids from systems of 1-truncated spaces. In the definable context, we decided to identify such spaces with internal sorts, we consider “local systems” of internal sorts, i.e., internal sorts of an expanded theory.

Our main result, Theorem 3.3.9, shows one direction of this expected correspondence: we associate to a 2-groupoid $\boldsymbol{G}$ in the theory $\mathcal{T}$ a theory $\mathcal{T}_{\boldsymbol{G}}$ expanding it, and a collection of internal sorts of $\mathcal{T}_{\boldsymbol{G}}$, which we view as “higher local systems”. The statement is that the canonical 2-groupoid associated to this datum recovers (up to weak equivalence) the original one (part of the other direction is indicated briefly, but is mostly left for future work).

We mention that this result is one possible generalisation of the results of [4] to higher dimensions. Other such generalisations include the papers [1; 2; 12], but they all appear to go in different directions. We also mention that in the context of usual (rather than definable) homotopy theory, analogous results are well known (for example, the main part of Theorem 3.3.9 is really a version of the higher-dimensional Yoneda lemma), but the methods in the proof of these results do not translate easily to the definable setting. In fact, the situation here is more similar to the one described in [8, §6.5], though made simpler by the existence of models (i.e., we have "enough points").

1.1. ***Structure of the paper.*** It is very simple: in Section 2, we review the situation in the one-dimensional case. This serves both as a motivating analogy and to complete some background used later. Most of the material there appears in some form in [4] (sometimes implicitly), but we include a few easy remarks regarding morphisms and equivalence, interpretation in terms of "local systems", and a different description of the groupoid associated to an internal cover (which already appeared slightly differently in [7]).

In Section 3, we expose the higher-dimensional picture, concentrating on dimension 2. We first define our higher internal covers, then review the theory of (truncated) Kan complexes and n-categories, with a few remarks special to the definable setting, and then prove the main result mentioned above (Theorem 3.3.9).

1.2. ***Conventions and terminology.*** For simplicity, we assume our theories $\mathcal{T}$ to admit elimination of quantifiers. By a *$\mathcal{T}$-structure* we mean a substructure of some model of $\mathcal{T}$. If A is such a $\mathcal{T}$-structure, by $\mathcal{T}_A$ we mean the expansion of $\mathcal{T}$ by constants for the elements of A, along with the usual axioms describing A. If A was not mentioned, we mean "for some A".

We also assume $\mathcal{T}$ eliminates imaginaries (this could be included in the general treatment, but would complicate the exposition). Our usage of elimination of imaginaries is often in the (equivalent) form of the existence of internal Homs: for every two definable sets $\boldsymbol{X}$ and $\boldsymbol{Y}$, there are an ind-definable set $\underline{\mathbf{Hom}}(\boldsymbol{X}, \boldsymbol{Y})$ and map $\mathbf{ev} : \boldsymbol{X} \times \underline{\mathbf{Hom}}(\boldsymbol{X}, \boldsymbol{Y}) \to \boldsymbol{Y}$, identifying the A-points of $\underline{\mathbf{Hom}}(\boldsymbol{X}, \boldsymbol{Y})$, for each $\mathcal{T}$-structure A, with the set of A-definable maps from $\boldsymbol{X}$ to $\boldsymbol{Y}$. It follows that the subset $\underline{\mathbf{Iso}}(\boldsymbol{X}, \boldsymbol{Y})$ of definable isomorphisms is also ind-definable.

Finally, we assume that each theory is generated by one sort, and finitely many relations. Similar to the case in [4], it can be seen that this assumption is not restrictive, since all our constructions commute with adding structure.

We recall that an *interpretation* of a theory $\mathcal{T}_1$ in another theory $\mathcal{T}_2$ is a model of $\mathcal{T}_1$ in the definable sets of $\mathcal{T}_2$: it assigns definable sets to the elements of the signature of $\mathcal{T}_1$, so that the axioms in $\mathcal{T}_1$ hold (this is often called a *definition* in the literature, which is equivalent to an interpretation under our assumption of elimination of imaginaries in $\mathcal{T}_2$). If $i : \mathcal{T}_1 \to \mathcal{T}_2$ is such an interpretation, it thus assigns to each definable set $\boldsymbol{X}$ of $\mathcal{T}_1$ a definable set $i(\boldsymbol{X})$ of $\mathcal{T}_2$. Since i is a model of $\mathcal{T}_1$, it assigns definable functions of $\mathcal{T}_2$ to definable functions of $\mathcal{T}_1$, and composition to composition, and thus determines a functor from the category $\mathcal{D}ef(\mathcal{T}_1)$ of definable sets of $\mathcal{T}_1$ to $\mathcal{D}ef(\mathcal{T}_2)$. We normally identify i with this functor, writing for example

$i(\boldsymbol{X})$ for the interpretation of a definable set $\boldsymbol{X}$ of $\mathcal{T}_1$. We note that an expansion is a particular case of an interpretation. We remark that not every functor from $\mathcal{D}ef(\mathcal{T}_1)$ to $\mathcal{D}ef(\mathcal{T}_2)$ arises from an interpretation: For example, an interpretation preserves all finite (inverse) limits (which always exist in $\mathcal{D}ef(\mathcal{T}_1)$). This is the main property of such functors that we use in this paper. A detailed description of categories of the form $\mathcal{D}ef(\mathcal{T})$ and functors that arise from interpretation occurs in [10], but we do not require it.

Similarly, if $i, j : \mathcal{T}_1 \to \mathcal{T}_2$ are interpretations, a map from i to j is an elementary map of models (equivalently a homomorphism, by our assumption of quantifier elimination), given by definable maps in $\mathcal{T}_2$. Equivalently, this is a natural transformation of functors, when i and j are viewed in this way. Such a map is an isomorphism if it has an inverse. An interpretation is called a *bi-interpretation* if there is an interpretation in the other direction such that both compositions are isomorphic to the identity.

When $\mathcal{T}_1$ and $\mathcal{T}_2$ are given with fixed interpretations $i_k : \mathcal{T} \to \mathcal{T}_k$ of a theory $\mathcal{T}$, we have versions of these notions over $\mathcal{T}$: an interpretation $j : \mathcal{T}_1 \to \mathcal{T}_2$ is over $\mathcal{T}$ if $j \circ i_1 = i_2$, and given two such interpretations $j_1, j_2 : \mathcal{T}_1 \to \mathcal{T}_2$, a map $\alpha : j_1 \to j_2$ is over $\mathcal{T}$ if $\alpha_{i_1(\boldsymbol{X})} : j_1(i_1(\boldsymbol{X})) \to j_2(i_1(\boldsymbol{X}))$ is the identity for all definable sets $\boldsymbol{X}$ of $\mathcal{T}$ (more naturally, we could require a given isomorphism from $j \circ i_1$ to i_2, but in practice we can always assume it to be the identity, and do that to simplify notation).

2. A review of the classical theory

2.1. ***Stable embeddings and internal covers.*** We recall the following classical definition of internal covers:

Definition 2.1.1. An expansion $\mathcal{T}^*$ of a theory $\mathcal{T}$ is an *internal cover* if $\mathcal{T}$ is stably embedded in $\mathcal{T}^*$, and for some expansion $\mathcal{T}^*_A$ of $\mathcal{T}^*$ by a set of constants A, each definable set in $\mathcal{T}^*_A$ is definably isomorphic to a definable set in $\mathcal{T}_{A_0}$, for some set of parameters A_0.

We recall that *stably embedded* here means that for every definable set $\boldsymbol{X}$ in $\mathcal{T}$, every subset of $\boldsymbol{X}$ definable in $\mathcal{T}^*$ with parameters from $\mathcal{T}^*$ is definable in $\mathcal{T}$, with parameters from $\mathcal{T}$.

It was noted in [6] that this condition can be reformulated as follows: if $i : \mathcal{T} \to \mathcal{T}^*$ is an expansion and $\boldsymbol{X}, \boldsymbol{Y}$ are definable in $\mathcal{T}$, there is a natural ($\mathcal{T}^*$-definable) map $i(\underline{\mathbf{Hom}}_{\mathcal{T}}(\boldsymbol{X}, \boldsymbol{Y})) \to \underline{\mathbf{Hom}}_{\mathcal{T}^*}(i(\boldsymbol{X}), i(\boldsymbol{Y}))$, and i is stably embedded precisely if this map is a bijection (for all $\boldsymbol{X}, \boldsymbol{Y}$ definable in $\mathcal{T}$). We note that in this case, the restriction of this map to the subset $\underline{\mathbf{Iso}}(\boldsymbol{X}, \boldsymbol{Y})$ of isomorphisms is also a bijection, and that taking into account parameters, no new structure is induced on $\mathcal{T}^*$. In particular, for any $\mathcal{T}$-structure A, the expansion $\mathcal{T}^*{}_A$ is well defined.

The same definition can be applied to a more general interpretation, so we say that an interpretation $i : \mathcal{T}_1 \to \mathcal{T}_2$ is *stable* if for all definable sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ of $\mathcal{T}_1$, the natural $\mathcal{T}_2$-definable map

$$i(\underline{\mathbf{Hom}}_{\mathcal{T}_1}(\boldsymbol{X}, \boldsymbol{Y})) \to \underline{\mathbf{Hom}}_{\mathcal{T}_2}(i(\boldsymbol{X}), i(\boldsymbol{Y}))$$

is a bijection. Explicitly, this means that each map definable with parameters from $i(\boldsymbol{X})$ to $i(\boldsymbol{Y})$ (in $\mathcal{T}_2$) "comes from" a unique map definable with parameters from $\boldsymbol{X}$ to $\boldsymbol{Y}$ (in $\mathcal{T}_1$). If i is viewed as a (left-exact) functor, as in Section 1.2, this is often stated as saying that i is Cartesian closed. In these terms, the definition of internal covers can be reformulated as follows:

Proposition 2.1.2. *An expansion $\mathcal{T}^*$ of a theory $\mathcal{T}$ is an internal cover if it is stable, and $\mathcal{T}^*$ admits a stable interpretation p in $\mathcal{T}_A$ over $\mathcal{T}$.*

As mentioned in Section 1.2, by "over $\mathcal{T}$" we mean that the restriction of p to $\mathcal{T}$ coincides with the expansion by constants.

Proof. Let $\boldsymbol{Q}$ be a definable set in $\mathcal{T}^*$, generating it over $\mathcal{T}$. Assume first that $\mathcal{T}^*$ is an internal cover of $\mathcal{T}$, so there is a sort $\boldsymbol{X}$ of $\mathcal{T}^*$, an expansion by a constant symbol $a \in \boldsymbol{X}$, and a definable bijection $g_a : \boldsymbol{Q} \to \boldsymbol{Q}_a$, with $\boldsymbol{Q}_a$ definable in $\mathcal{T}$. By stable-embeddedness, $\boldsymbol{Q}_a$ is definable by a parameter a_0 in $\mathcal{T}$. The assignment $\boldsymbol{Q} \mapsto \boldsymbol{Q}_a$ extends uniquely to an interpretation x_{a_0} of $\mathcal{T}^*$ in $\mathcal{T}_{a_0}$, over $\mathcal{T}$. Since g_a determines a definable isomorphism between $\boldsymbol{Q}$ and $\boldsymbol{Q}_a$ (and similarly for any definable set it generates), this interpretation is stable.

Conversely, assume we have a stable interpretation $p : \mathcal{T}^* \to \mathcal{T}_{A_0}$ over i. We still denote by i the extension of i to the expansion $\mathcal{T}_{A_0} \to \mathcal{T}^*_{A_0}$ which is the identity on A_0 (it is still stable). Setting $\boldsymbol{C} = p(\boldsymbol{Q})$, the set $\underline{\mathbf{Iso}}(p(\boldsymbol{Q}), p(i(\boldsymbol{C}))) = \underline{\mathbf{Iso}}(\boldsymbol{C}, \boldsymbol{C})$ is nonempty, since it contains the identity on $\boldsymbol{C}$. Since p is stable, the left-hand side admits a definable bijection with $\underline{\mathbf{Iso}}(\boldsymbol{Q}, i(\boldsymbol{C}))$, so is nonempty as well. Any point a of this set shows that $\mathcal{T}^*$ is an internal cover. $\square$

2.2. *Definable groupoids.* A definable groupoid is denoted as $\boldsymbol{G} = \langle \boldsymbol{G}_0, \boldsymbol{G}_1 \rangle$, with a definable set $\boldsymbol{G}_0$ of objects and a definable set $\boldsymbol{G}_1$ of isomorphisms, where the domain and codomain maps are denoted $\boldsymbol{d}, \boldsymbol{c} : \boldsymbol{G}_1 \to \boldsymbol{G}_0$, respectively, and composition denoted by $\circ$. For objects $a, b \in \boldsymbol{G}_0$, we write $\boldsymbol{G}(a, b)$ for the a, b-definable set $\langle \boldsymbol{d}, \boldsymbol{c} \rangle^{-1}(\langle a, b \rangle)$ of morphisms from a to b. A map $f : \boldsymbol{G} \to \boldsymbol{H}$ of definable groupoids is a definable functor: a pair of maps $f_0 : \boldsymbol{G}_0 \to \boldsymbol{H}_0$ and $f_1 : \boldsymbol{G}_1 \to \boldsymbol{H}_1$ commuting with the domain, codomain and composition maps. We say that f is a *weak equivalence*, denoted $f : \boldsymbol{G} \xrightarrow{\sim} \boldsymbol{H}$, if it induces an equivalence of categories on all models (this terminology is generalised in Definition 3.2.7). Our groupoids are generally *not* assumed to be connected.

In [4, §2], a definable groupoid is attached to each internal cover. This groupoid also admits two descriptions. The first is as a definable groupoid $\boldsymbol{G}^*$ in $\mathcal{T}^*$; this construction depends on the choice of a definable set $\boldsymbol{X}$ in $\mathcal{T}^*$, as in the proof of Proposition 2.1.2. Given this choice, the groupoid can be described as follows:

Construction 2.2.1. The groupoid associated to the data above is described as follows:

- **Objects ($\boldsymbol{G}_0$):** Complete types of elements $a \in \boldsymbol{X}$ over $\mathcal{T}$, along with an additional object $*$. Since $\mathcal{T}$ is stably embedded, this set of types is definable in $\mathcal{T}$ (it is definable, rather than pro-definable, by our finiteness assumption on the language in Section 1.2).
- **Morphisms ($\boldsymbol{G}_1$):** The set of isomorphisms from $*$ to a type $p \in \boldsymbol{G}_0$ is given by the realisations of p. Given another type $q \in \boldsymbol{G}_0$, a morphism from p to q is given by a 2-type s extending p and q (over $\mathcal{T}$). Distinct realisations of s correspond to distinct ways of writing the morphism s as a composition of a morphism from p to $*$ and a morphism from $*$ to q.
- **Composition:** Given a type $s(x, y) \in \boldsymbol{G}_1$ extending $p(x), q(y) \in \boldsymbol{G}_0$, and a type $t(y, z) \in \boldsymbol{G}_1$ extending $q(y)$ and $r(z) \in \boldsymbol{G}_0$, the internality assumption implies that there is a unique 3-type $u(x, y, z)$ extending all of them. The restriction u to x, z is the composition of s, t.

  The composition of an isomorphism a from $*$ to $p \in \boldsymbol{G}_0$ with (the inverse of) another such isomorphism b to $q \in \boldsymbol{G}_0$ is the type of $\langle a, b\rangle$. The other compositions are determined by these conditions. □

We denote by $\boldsymbol{G}$ the full subgroupoid of $\boldsymbol{G}^*$ on the same objects excluding $*$. Then $\boldsymbol{G}$ is defined entirely in $\mathcal{T}$.

2.2.2. To give a second description, consider, for each $\mathcal{T}$-structure A, the groupoid $I(A) = I_{\mathcal{T}^*/\mathcal{T}}(A)$ whose objects are stable interpretations of $\mathcal{T}^*$ in $\mathcal{T}_A$, that are the identity on $\mathcal{T}_A$, and whose morphisms are isomorphisms of such interpretations, which are the identity when restricted to $\mathcal{T}$. Here again we may enlarge I to obtain I^*, by adding an additional object $*$, which is described explicitly as the identity interpretation of $\mathcal{T}^*$, and again morphisms are given by A^*-definable isomorphisms of interpretations over $\mathcal{T}$ (where A^* is now a $\mathcal{T}^*$-structure). The following statement appeared in a slightly different form in [7]:

Proposition 2.2.3. *With notation as in 2.2, for each $\mathcal{T}^*$-structure A, there is a fully faithful embedding $i_A : \boldsymbol{G}^*(A) \to \boldsymbol{I}^*(A)$, preserving the vertex, and commuting with action by automorphisms on A. If A is a model, i_A is an equivalence of categories.*

Proof. This is essentially [4, Theorem 3.2]. The functor i_A was described in the proof of Proposition 2.1.2: to an object p of $\boldsymbol{G}(A)$, viewed as a type over $\mathcal{T}$ (definable over A_0), we attach the interpretation $x_b = x_p$ described there, with b any realisation of p (as explained there, x_b depends only on $p \in A_0$ and not on b).

Each such realisation determines an isomorphism g_b from x_b to $*$, again as above, which describes the functor on morphisms from p to $*$. If q is another object, with realisation c, i_A assigns ${g_c}^{-1} \circ g_b : x_p \to x_q$ to the type r of the pair (b, c). This depends only on r, since the code for this composition lies in $\mathcal{T}$, by stable embeddedness. This code also determines r completely, so the functor is fully faithful.

To prove the final statement, let $i : \mathcal{T}^* \to \mathcal{T}$ be any interpretation over a model M_0. The internality assumption implies that for some $p \in \boldsymbol{G}(M_0)$, the set Y of isomorphisms between $*$ and p is nonempty. Since M_0 is a model, there is a point b in $i(Y)(M_0)$. Then g_b is an isomorphism from x_b to i. □

To summarise, to each stable embedding of $\mathcal{T}$ in $\mathcal{T}^*$, we had attached a groupoid $I_{\mathcal{T}^*/\mathcal{T}}$ classifying stable interpretations of $\mathcal{T}^*$ back in $\mathcal{T}$. The embedding is an internal cover precisely if the groupoid is nonempty, and in this case, the groupoid I is equivalent to a definable one (and to the classical binding groupoid). Conversely, starting with a definable groupoid $\boldsymbol{G}$ in $\mathcal{T}$, there is an internal cover $\mathcal{T}^* = \mathcal{T}_{\boldsymbol{G}}$ and an equivalence $\boldsymbol{G} \to I_{\mathcal{T}^*/\mathcal{T}}$:

Construction 2.2.4. The theory $\mathcal{T}_{\boldsymbol{G}}$ expands $\mathcal{T}$ by an additional sort $\boldsymbol{X}$, a function symbol $c : \boldsymbol{X} \to \boldsymbol{G}_0$, and a function symbol $a : \boldsymbol{X} \times_{\boldsymbol{G}_0} \boldsymbol{G}_1 \to \boldsymbol{X}$, where $\boldsymbol{X} \times_{\boldsymbol{G}_0} \boldsymbol{G}_1 = \{\langle x, g\rangle \mid c(x) = d(g)\}$, and $\mathcal{T}_{\boldsymbol{G}}$ states that $\boldsymbol{X}$ is nonempty, and that the resulting structure is a groupoid $\boldsymbol{G}^*$ extending $\boldsymbol{G}$ by an additional object $*$, with elements $x \in \boldsymbol{X}$ viewed as morphisms from $*$ to $c(x)$, and a provides the composition of such elements with morphisms of $\boldsymbol{G}$.

Starting from this $\mathcal{T}_{\boldsymbol{G}}$, each element of $\boldsymbol{X}$ exhibits $\mathcal{T}_{\boldsymbol{G}}$ as an internal cover of $\mathcal{T}$. The type of such an element $x \in \boldsymbol{X}$ over $\mathcal{T}$ is given by $c(x)$, and the realisations of this type are indeed the morphisms from $*$ to $c(x)$, so Construction 2.2.1 indeed recovers $\boldsymbol{G}$. □

For example, when $\boldsymbol{G}$ is a definable group (groupoid with one object), the resulting $\mathcal{T}_{\boldsymbol{G}}$ is the theory of $\boldsymbol{G}$-torsors. We refer to [4, §3] for more details, but note again that our construction is slightly different when $\boldsymbol{G}$ is not connected: we always expand just by one additional object $*$, thus obtaining an internal cover (possibly incomplete), even in the nonconnected case.

Alternatively, Definition 3.3.1 is a generalisation that also applies to this case.

2.2.5. *Definable $\boldsymbol{G}$-sets.* If $\boldsymbol{G} = \langle \boldsymbol{G}_0, \boldsymbol{G}_1\rangle$ is a definable groupoid in $\mathcal{T}$, by a (left) *$\boldsymbol{G}$-set* we mean a definable set $\boldsymbol{X}$, a definable map $\pi : \boldsymbol{X} \to \boldsymbol{G}_0$ to the set of objects $\boldsymbol{G}_0$ of $\boldsymbol{G}$, and an "action" map $a : \boldsymbol{G}_1 \times_{\boldsymbol{G}_0} \boldsymbol{X} \to \boldsymbol{X}$, over $\boldsymbol{G}_0$, satisfying the usual action axioms (here, $\boldsymbol{G}_1 \times \boldsymbol{G}_0 \boldsymbol{X}$ is the definable subset of $\boldsymbol{G}_1 \times \boldsymbol{X}$ given by $d(g) = \pi(x)$, and "over $\boldsymbol{G}_0$" means that $c(g) = \pi(a(g, x))$ for all such pairs). Thus, a morphism $g : a \to b$ in $\boldsymbol{G}$ determines a bijection $a_g : \boldsymbol{X}_a \to \boldsymbol{X}_b$, where $\boldsymbol{X}_t = \pi^{-1}(t)$, and we sometimes write gx in places of $a_g(x)$ (a pair $\langle \boldsymbol{G}, \boldsymbol{X}\rangle$ as above is called a *concrete* *groupoid* in [4, §3]). We think of $\boldsymbol{G}$-sets as analogues of local systems over $\boldsymbol{G}$. A morphism from a $\boldsymbol{G}$-set $\boldsymbol{X}$ to another $\boldsymbol{G}$-set $\boldsymbol{Y}$ is a definable map from $\boldsymbol{X}$ to $\boldsymbol{Y}$ that commutes with π and a.

Let $\boldsymbol{X}$ be a $\boldsymbol{G}$-set. If $\boldsymbol{H} = \langle \boldsymbol{H}_0, \boldsymbol{H}_1\rangle$ is another groupoid, and $i : \boldsymbol{G} \to \boldsymbol{H}$ is a definable map of groupoids, we set

$$i_!(\boldsymbol{X}) = \{\langle h, x\rangle \in \boldsymbol{H}_1 \times \boldsymbol{X} \mid i(\pi(x)) = \boldsymbol{d}(h)\}/\sim,$$

where $\langle h, gx\rangle \sim \langle h \circ i(g), x\rangle$ for $g \in \boldsymbol{G}_1$ satisfying $\boldsymbol{d}(g) = \pi(x)$ and $i(\boldsymbol{c}(g)) = \boldsymbol{d}(h)$. This is an $\boldsymbol{H}$-set, with structure map induced by $\langle h, x\rangle \mapsto \boldsymbol{c}(h)$ and action induced by $\langle h', \langle h, x\rangle\rangle \mapsto \langle h'h, x\rangle$. On the other hand, if $\boldsymbol{Y}$ is an $\boldsymbol{H}$-set, we set $i^*(\boldsymbol{Y}) = \boldsymbol{G}_0 \times_{\boldsymbol{H}_0} \boldsymbol{Y}$, with the projection to $\boldsymbol{G}_0$ as the structure map, and action given by $\langle g, y\rangle \mapsto i(g)y$ for $y \in \boldsymbol{Y}$ with $\pi(y) = i(\boldsymbol{d}(g))$. It is clear that both constructions are functorial, and as the notation suggests, $i_!$ is left adjoint to i^*.

With these notions, we have the following description of definable sets in $\mathcal{T}^*$ as local systems over $\boldsymbol{G}$:

Proposition 2.2.6. *If $\mathcal{T}^*$ is an internal cover of $\mathcal{T}$, corresponding to the definable groupoid $\boldsymbol{G}$ in $\mathcal{T}$, then the category of definable sets in $\mathcal{T}^*$ is equivalent to the category of $\boldsymbol{G}$-sets in $\mathcal{T}$. Definable sets from $\mathcal{T}$ correspond to themselves, with trivial action.*

Proof. To each definable set $\boldsymbol{X}^*$ in $\mathcal{T}^*$ we assign the definable set $\boldsymbol{X} = \coprod_{p \in \boldsymbol{G}_0} p(\boldsymbol{X}^*)$. It follows from the uniformity of p that $\boldsymbol{X}$ is definable in $\mathcal{T}$. By definition, $\boldsymbol{X}$ admits a definable map to $\boldsymbol{G}_0$. The action is given tautologically by the identification of the morphisms in $\boldsymbol{G}_0$ with maps of interpretations. Since each p is an interpretation, this is functorial in $\boldsymbol{X}^*$.

In the other direction, let $\boldsymbol{G}^*$ be the canonical extension of $\boldsymbol{G}$ in $\mathcal{T}^*$ (we identify $\boldsymbol{G}$ with its image in $\mathcal{T}^*$), let $i : \boldsymbol{G} \to \boldsymbol{G}^*$ be the inclusion, and let j be the inclusion of the canonical object $*$ of $\boldsymbol{G}^*$, along with its automorphism group $\boldsymbol{H}$, into $\boldsymbol{G}^*$. A definable $\boldsymbol{G}$-set $\boldsymbol{X}$ in $\mathcal{T}$, viewed again as embedded in $\mathcal{T}^*$, corresponds then to $\boldsymbol{X}^* = j^*(i_!(\boldsymbol{X}))$ (and the resulting action by $\boldsymbol{H}$ is the natural action by automorphisms). □

We note that each definable set in $\mathcal{T}^*$ comes equipped with an action of the binding group $\underline{\mathbf{Aut}}(*)$, and with it, the first direction could likewise be described as $\boldsymbol{X} = i^*(j_!(\boldsymbol{X}^*))$.

Corollary 2.2.7. *If $\boldsymbol{G}$ is a groupoid associated to an internal cover $\mathcal{T}^*$ of $\mathcal{T}$, then $\mathcal{T}^*$ is bi-interpretable with $\mathcal{T}_{\boldsymbol{G}}$ over $\mathcal{T}$.*

Proof. The definable groupoid $\boldsymbol{G}^*$ in $\mathcal{T}^*$ forms an interpretation of $\mathcal{T}_{\boldsymbol{G}}$ over $\mathcal{T}$. It is a bi-interpretation since both categories of definable sets are equivalent to the category of $\boldsymbol{G}$-sets in $\mathcal{T}$ (commuting with the above interpretation). □

2.2.8. *Pushouts.* Let $g : \boldsymbol{K} \to \boldsymbol{G}$ and $h : \boldsymbol{K} \to \boldsymbol{H}$ be maps of definable groupoids, and assume that g is fully faithful. We construct another definable groupoid $\boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H}$ that can be viewed as the pushout of $\boldsymbol{G}$ and $\boldsymbol{H}$ over $\boldsymbol{K}$ as follows: For objects, we let $(\boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H})_0 = \boldsymbol{G}_0 \amalg \boldsymbol{H}_0$. If a, b are two such objects, we define the morphisms as follows:

(1) If $a, b \in \boldsymbol{H}_0$, then $(\boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H})(a, b) = \boldsymbol{H}(a, b)$.

(2) If $a \in \boldsymbol{G}_0$ and $b \in \boldsymbol{H}_0$, morphisms from a to b are equivalence classes $v \otimes u$ of pairs $\langle v, u\rangle$, where $u \in \boldsymbol{G}(a, g(c))$, $v \in \boldsymbol{H}(h(c), b)$ for some $c \in \boldsymbol{K}_0$, and $\langle v, g(w) \circ u\rangle$ is equivalent to $\langle v \circ h(w), u\rangle$ for all $w \in \boldsymbol{K}_1$ for which the composition is defined. Morphisms from b to a are defined analogously.

(3) If $a, b \in \boldsymbol{G}_0$ are both in the essential image of g, a morphism from a to b is similarly defined as an equivalence class $u' \otimes v \otimes u$, with $u, u' \in \boldsymbol{G}_1$ and $v \in \boldsymbol{H}_1$.

(4) If either of $a, b \in \boldsymbol{G}_0$ is not in the essential image of g, then morphisms are the same as in $\boldsymbol{G}$.

The composition $(u' \otimes v \otimes u) \circ (u'_1 \otimes v_1 \otimes u_1)$ is defined as follows: There are $a, b \in \boldsymbol{K}_0$ such that $u \circ u'_1$ is a morphism from $g(a)$ to $g(b)$. Since g is fully faithful, it has the form $g(w)$ for a unique morphism w from a to b in $\boldsymbol{K}$. We define the composition to be $u' \otimes (v \circ h(w) \circ v_1) \otimes u_1$. It is clear that this is independent of the choices of representatives. The composition in the other cases is defined similarly.

There is an obvious map $h' : \boldsymbol{H} \to \boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H}$, and we define $g' : \boldsymbol{G} \to \boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H}$ by sending each object to itself, each morphism between objects not in the essential image of g to itself as well, and for $a, b \in \boldsymbol{G}_0$ in the essential image of g, and u a morphism from a to b, we set $g'(u) = (u \circ u'^{-1}) \otimes \mathbf{1}_{h(c)} \otimes u'$, where $u' : a \to g(c)$ is any morphism and $c \in \boldsymbol{K}_0$. We have an isomorphism α from $h' \circ h$ to $g' \circ g$, given on an object $c \in \boldsymbol{K}_0$ by $\mathbf{1}_{g(c)} \otimes \mathbf{1}_{h(c)}$. It is routine to check that everything is well defined, and also that the following statement holds.

Proposition 2.2.9. *Let $g : \boldsymbol{K} \to \boldsymbol{G}$, $h : \boldsymbol{K} \to \boldsymbol{H}$ and the rest of the notation be as above.*

(1) *Given definable maps of groupoids $g_1 : \boldsymbol{G} \to \boldsymbol{F}$ and $h_1 : \boldsymbol{H} \to \boldsymbol{F}$, and an isomorphism $\beta : h_1 \circ h \to g_1 \circ g$, there is a unique map of groupoids $f : \boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H} \to \boldsymbol{F}$ that coincides with g_1 and h_1 on the objects, $f \circ g' = g_1$, $f \circ h' = h_1$ and such that $f \cdot \alpha = \beta$.*

(2) *$h' : \boldsymbol{H} \to \boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H}$ is fully faithful. If g is a weak equivalence, then so is h'.*

Remark 2.2.10. We could make a similar construction where the set of objects is $\boldsymbol{G}_0 \amalg_{\boldsymbol{K}_0} \boldsymbol{H}_0$ in place of the disjoint union, and with α the identity. The last proposition provides a map from $\boldsymbol{G} \otimes_{\boldsymbol{K}} \boldsymbol{H}$ to this variant, which is easily seen to be a weak equivalence. We use the two constructions interchangeably. □

Remark 2.2.11. Without the assumption that one of the maps is fully faithful, the pushout need not be definable. For example, when all groupoids are groups, this is the usual free product with amalgamation. □

2.3. *Maps of groupoids and of interpretations.* With stable interpretations over $\mathcal{T}$, the assignment $\mathcal{T}^* \mapsto I_{\mathcal{T}^*/\mathcal{T}}$ is contravariantly functorial in $\mathcal{T}^*$, and fully faithful: a stable interpretation $i : \mathcal{T}_1 \to \mathcal{T}_2$ over $\mathcal{T}$ induces a functor $i^* : I_{\mathcal{T}_2/\mathcal{T}} \to I_{\mathcal{T}_1/\mathcal{T}}$ by composition.

In the other direction, if $f : \boldsymbol{G} \to \boldsymbol{H}$ is a map of definable groupoids, corresponding to internal covers $\mathcal{T}_{\boldsymbol{G}}$ and $\mathcal{T}_{\boldsymbol{H}}$, f determines a stable interpretation $i^f : \mathcal{T}_{\boldsymbol{H}} \to \mathcal{T}_{\boldsymbol{G}}$ over $\mathcal{T}$ that can be described in at least two ways:

(1) An interpretation of $\mathcal{T}_{\boldsymbol{H}}$ over $\mathcal{T}$ is determined by its value on the extended groupoid $\boldsymbol{H}^*$ defined in $\mathcal{T}_{\boldsymbol{H}}$. We set $i^f(\boldsymbol{H}^*) = \boldsymbol{G}^* \otimes_{\boldsymbol{G}} \boldsymbol{H}$ (with respect to the given map f). This makes sense since the inclusion of $\boldsymbol{G}$ in $\boldsymbol{G}^*$ is a weak equivalence, and is an interpretation since the embedding of $\boldsymbol{H}$ in $\boldsymbol{G}^* \otimes_{\boldsymbol{G}} \boldsymbol{H}$ is a weak equivalence that misses precisely one object $*$, and this completely determines its theory. To see that it is stable, we may first choose a parameter in $\boldsymbol{G}^*$. But then i^f is identified with one of the standard interpretations into $\mathcal{T}$.

(2) Alternatively, we may use Proposition 2.2.6 to identify definable sets in $\mathcal{T}_{\boldsymbol{G}}$ and in $\mathcal{T}_{\boldsymbol{H}}$ with $\boldsymbol{G}$- and $\boldsymbol{H}$-sets in $\mathcal{T}$. Then i^f is identified with f^* (in this approach, it is less direct to see that one gets a stable interpretation).

It is easy to verify that $(i^f)^* = f$ (after identifying $\boldsymbol{G}$ with its image in $I_{\mathcal{T}_G/\mathcal{T}}$ via Proposition 2.2.3, and similarly for $\boldsymbol{H}$). However, not every stable interpretation $i : \mathcal{T}_{\boldsymbol{H}} \to \mathcal{T}_{\boldsymbol{G}}$ (over $\mathcal{T}$) is of the form i^f for some $f : \boldsymbol{G} \to \boldsymbol{H}$. The other source of interpretations comes from the other operation described in 2.2.5: when f is a weak equivalence, the composition of f with the inclusion of $\boldsymbol{H}$ in $\boldsymbol{H}^*$ is a weak equivalence, so restricting to the image of f (on the objects), we obtain an interpretation i_f of $\boldsymbol{G}^*$ (hence of $\mathcal{T}_{\boldsymbol{G}}$).

Proposition 2.3.1. *Let $\boldsymbol{G}$ and $\boldsymbol{H}$ be two definable groupoids, with associated covers $\mathcal{T}_{\boldsymbol{G}}$ and $\mathcal{T}_{\boldsymbol{H}}$. Then every stable interpretation $i : \mathcal{T}_{\boldsymbol{G}} \to \mathcal{T}_{\boldsymbol{H}}$ over $\mathcal{T}$ is obtained as a composition $i = i^f \circ i_g$, for some definable groupoid $\boldsymbol{K}$, definable map $f : \boldsymbol{H} \to \boldsymbol{K}$ and weak equivalence $g : \boldsymbol{G} \xrightarrow{\sim} \boldsymbol{K}$.*

In particular, if $i : \mathcal{T}_1 \to \mathcal{T}_2$ is a stable interpretation of internal covers over $\mathcal{T}$, we may choose definable groupoids $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ corresponding to the $\mathcal{T}_i$, so that i is induced by a map $f : \boldsymbol{G}_2 \to \boldsymbol{G}_1$ of groupoids (up to bi-interpretation).

A configuration of the form $\langle \boldsymbol{K}, f, g \rangle$ as above is called a *cospan* from $\boldsymbol{H}$ to $\boldsymbol{G}$.

Proof. $\boldsymbol{H}$ embeds in $I_{\mathcal{T}_H/\mathcal{T}}$ via Proposition 2.2.3, which maps via i^* to $I_{\mathcal{T}_G/\mathcal{T}}$. We set $f : \boldsymbol{H} \to \boldsymbol{K}$ to be the restriction of i^* to $\boldsymbol{H}$, where $\boldsymbol{K}$ denotes any definable weakly equivalent subgroupoid of $I_{\mathcal{T}_G/\mathcal{T}}$, which also contains $\boldsymbol{G}$. Then g is the inclusion of $\boldsymbol{G}$ in $\boldsymbol{K}$.

The last part follows (using Corollary 2.2.7) by choosing $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ arbitrarily, and then replacing $\boldsymbol{G}_1$ by $\boldsymbol{K}$ as above. $\square$

As in the construction of the pushout, we may choose $\boldsymbol{K}$ so that its objects are the disjoint union of the objects of $\boldsymbol{G}$ and $\boldsymbol{H}$, and we always assume that this is the case. In the case when i is a bi-interpretation, we recover the notion of equivalence from [4, §3].

2.3.2. *Composition and isomorphisms.* Assume that for groupoids $\boldsymbol{F}$, $\boldsymbol{G}$ and $\boldsymbol{H}$ in $\mathcal{T}$, we are given interpretations $i : \mathcal{T}_{\boldsymbol{F}} \to \mathcal{T}_{\boldsymbol{G}}$ and $j : \mathcal{T}_{\boldsymbol{G}} \to \mathcal{T}_{\boldsymbol{H}}$, represented by cospans $g_1 : \boldsymbol{F} \overset{\sim}{\dashrightarrow} \boldsymbol{K}_1$, $f_1 : \boldsymbol{G} \to \boldsymbol{K}_1$, $g_2 : \boldsymbol{G} \overset{\sim}{\dashrightarrow} \boldsymbol{K}_2$ and $f_2 : \boldsymbol{H} \to \boldsymbol{K}_2$ as in Proposition 2.3.1. Since g_2 is a weak equivalence, we may form the pushout $\boldsymbol{K} = \boldsymbol{K}_1 \otimes_{\boldsymbol{G}} \boldsymbol{K}_2$. By Proposition 2.2.9, the map from $\boldsymbol{K}_1$ to $\boldsymbol{K}$ is a weak equivalence, and therefore so is the composed map g. Hence, g along with the composed map $f : \boldsymbol{H} \to \boldsymbol{K}$ form a cospan that represents a stable interpretation of $\mathcal{T}_{\boldsymbol{F}}$ in $\mathcal{T}_{\boldsymbol{H}}$. To conform with the decision about the objects of the representing groupoid $\boldsymbol{K}$, we remove the intermediate two copies of $\boldsymbol{G}_0$, and denote the resulting groupoid by $\boldsymbol{K}_2 \circ \boldsymbol{K}_1 = \boldsymbol{K}_2 \circ_{\boldsymbol{G}} \boldsymbol{K}_1$ (though it does depend on the additional data). The following is a direct calculation.

Proposition 2.3.3. *In the above situation, the maps $g : \boldsymbol{F} \overset{\sim}{\dashrightarrow} \boldsymbol{K}_2 \circ \boldsymbol{K}_1$ and $f : \boldsymbol{H} \to \boldsymbol{K}_2 \circ \boldsymbol{K}_1$ represent the composed interpretation $j \circ i$.*

Finally, we consider isomorphisms of interpretations between (stable) interpretations of internal covers over $\mathcal{T}$.

Proposition 2.3.4. *Let $i, j : \mathcal{T}_{\boldsymbol{G}_1} \to \mathcal{T}_{\boldsymbol{G}_2}$ be two stable interpretations of internal covers over $\mathcal{T}$. Assume i is represented by a cospan $i_1 : \boldsymbol{G}_1 \overset{\sim}{\dashrightarrow} \boldsymbol{H}_1$ and $i_2 : \boldsymbol{G}_2 \to \boldsymbol{H}_1$, and j by $j_1 : \boldsymbol{G}_1 \overset{\sim}{\dashrightarrow} \boldsymbol{H}_2$, $j_2 : \boldsymbol{G}_2 \to \boldsymbol{H}_2$, with each set of objects of $\boldsymbol{H}_n$ the disjoint union of the objects of $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ (realised by the object parts of i_k and j_k).*

Then there is a natural bijection between isomorphisms $\alpha : i \to j$ (over $\mathcal{T}$) and isomorphisms $\widetilde{\alpha} : \boldsymbol{H}_1 \to \boldsymbol{H}_2$ which are the identity on the images of $\boldsymbol{G}_1$, $\boldsymbol{G}_2$.

As an example, if $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are groups, then each $\boldsymbol{H}_i$ corresponds to a $\boldsymbol{G}_1 - \boldsymbol{G}_2$ bitorsor, and an isomorphism of the corresponding interpretations corresponds to an isomorphism of such bitorsors.

Proof. Let $\boldsymbol{P}_l$ be the set of arrows in $\boldsymbol{H}_l$ with domain in $\boldsymbol{G}_1$ and codomain in $\boldsymbol{G}_2$. This is a $\boldsymbol{G}_1$-set, with structure given by the domain map and composition. The interpretation i takes $\boldsymbol{P}_1$ to the $\boldsymbol{G}_2$-set given by viewing the arrows in $\boldsymbol{P}_1$ in the other direction, and likewise with j and $\boldsymbol{P}_2$. So the map α maps $\boldsymbol{P}_1$ to $\boldsymbol{P}_2$, compatibly with the composition. This is the same as giving an isomorphism $\widetilde{\alpha}$ as in the statement.

The rest of the structure is induced by the $\boldsymbol{P}_i$, so α is determined by $\widetilde{\alpha}$. Conversely, each $\widetilde{\alpha}$ as in the statement extends to an interpretation. $\square$

We summarise most of the content of this section in the following theorem (mostly contained in [4, §3]):

Theorem 2.3.5. *Let $\mathcal{T}$ be a theory. Each internal cover $\mathcal{T}^*$ of $\mathcal{T}$ is bi-interpretable over $\mathcal{T}$ with an internal cover of the form $\mathcal{T}_{\boldsymbol{G}}$. An interpretation of $\mathcal{T}_{\boldsymbol{H}}$ in $\mathcal{T}_{\boldsymbol{G}}$ corresponds to a cospan from $\boldsymbol{G}$ to $\boldsymbol{H}$, and each such cospan determines an interpretation. Maps between interpretations correspond to maps between cospans.*

In particular, covers $\mathcal{T}_1$ and $\mathcal{T}_2$ are bi-interpretable over $\mathcal{T}$ if and only if the corresponding groupoids are equivalent.

More succinctly (and slightly more precisely), the bicategory of internal covers over $\mathcal{T}$ is equivalent to the bicategory of definable groupoids in $\mathcal{T}$ (with morphisms given by cospans and morphisms between them given by bitorsors). See also Remark 3.3.4.

Proof. This is a combination of Corollary 2.2.7 with Propositions 2.3.1 and 2.3.4. $\square$

The description above exhibits the groupoid associated to an expansion as interpretations of $\mathcal{T}^*$ in $\mathcal{T}$. In [4], it was suggested that definable sets of an internal cover of $\mathcal{T}$ can be viewed as generalised imaginary sorts of $\mathcal{T}$. With this point of view, it is natural to ask for the structure classifying interpretations of such sorts as well. However, such generalised sorts have more structure: in addition to the sorts themselves and maps between them (interpretations), we also have maps between maps. The notion of equivalence should be modified as well: it is no longer reasonable to expect a bijection on the level of morphisms. In fact, as the 1-dimensional case already shows, it is not reasonable to expect even a map.

3. Generalised imaginaries

We now suggest how internal covers can play the role of definable sets in the above description, by going one dimension higher.

3.1. *Higher internal covers.*

Definition 3.1.1. Let $\mathcal{T}$ be a theory, $\mathcal{T}_1$ and $\mathcal{T}_2$ internal covers of $\mathcal{T}$. For every set of parameters A for $\mathcal{T}$, we denote by $\mathrm{Hom}_{\mathcal{T}}(\mathcal{T}_1, \mathcal{T}_2)(A)$ the groupoid whose objects are stable interpretations of $\mathcal{T}_1$ in $\mathcal{T}_{2A}$, over $\mathcal{T}_A$, and whose morphisms are isomorphisms of interpretations over $\mathcal{T}$.

Thus, what we denoted by I above is $\mathrm{Hom}_{\mathcal{T}}(\mathcal{T}^*, \mathcal{T})$. Similar to that case, $\mathrm{Hom}_{\mathcal{T}}(\mathcal{T}_1, \mathcal{T}_2)$ is ind-definable in $\mathcal{T}$: if $\mathcal{T}_i = \mathcal{T}_{\boldsymbol{G}_i}$ for $\mathcal{T}$-definable groupoids $\boldsymbol{G}_1$, $\boldsymbol{G}_2$, each interpretation above can be described like in Section 2.3 as given by certain definable maps $\boldsymbol{G}_i \to \boldsymbol{H}$, a definable condition. Similarly, isomorphisms between interpretations are given by the $\mathcal{T}$-definable families of maps as in Proposition 2.3.4, uniform in the $\boldsymbol{H}_i$ (see Section 3.3.6 for a more detailed description).

An interpretation between theories extends to internal sorts: if $i : \mathcal{T} \to \mathcal{S}$ is an interpretation, and $\widetilde{\mathcal{T}}$ is an internal cover of $\mathcal{T}$ associated to the groupoid $\boldsymbol{G}$ in $\mathcal{T}$, we denote by $i(\widetilde{\mathcal{T}})$ the internal cover of $\mathcal{S}$ associated to $i(\boldsymbol{G})$.

We now wish to define (slightly) higher analogues of stable embeddings and internal covers. One discrepancy with the 1-dimensional case occurs as follows: If $\mathcal{T}$ is an internal cover of $\mathcal{T}_0$, we might be interested in only part of the structure on $\mathcal{T}$ when considering, for example, the Galois group. As long as this partial structure includes the definable sets witnessing the internality, this can be done be replacing $\mathcal{T}$ with a reduct including only those definable sets. In the higher version, definable sets are replaced by definable groupoids in $\mathcal{T}$ (equivalently, internal covers), and again we may wish to restrict to a partial collection. However, there is no reason to expect that this partial collection is the full collection of definable groupoids in some reduct of $\mathcal{T}$. Furthermore, the internality condition for 0-definable sets automatically implies it for definable sets over parameters. Again, there is no reason to expect a similar statement for groupoids. For this reason, our definition depends on the auxiliary data Γ consisting of families of definable groupoids, which are the groupoids we wish to preserve. More precisely, we have the following.

Definition 3.1.2. Let $\mathcal{T}$ be a theory. The data of *distinguished covers* for $\mathcal{T}$ consists of the following:

(1) An ind-definable family Γ_0 of internal covers of $\mathcal{T}$ (equivalently, of definable groupoids in $\mathcal{T}$).

(2) An ind-definable family of interpretations over $\mathcal{T}$ between any two covers $\mathcal{T}_1, \mathcal{T}_2 \in \Gamma_0$, depending definably on $\mathcal{T}_1, \mathcal{T}_2$ and closed under composition (the full definable family is denoted by Γ_1).

(3) For every two interpretations $f, g : \mathcal{T}_1 \to \mathcal{T}_2$ in Γ_1, an ind-definable family of isomorphisms from f to g, closed under composition and restricting to the identity on $\mathcal{T}$. Again we assume that the family of all such isomorphisms is uniformly (ind-)definable in f, g, and denote it by Γ_2.

If $\mathcal{T}_0$ is a reduct of $\mathcal{T}$, we say that $\Gamma = \langle \Gamma_0, \Gamma_1, \Gamma_2 \rangle$ is over $\mathcal{T}_0$ if the parameters for the ind-definable families above range over definable sets in $\mathcal{T}_0$.

We note that in terms of definable groupoids, the closure under composition translates to closure under the composition operation from Section 2.3.2.

If a theory $\mathcal{T}$ is given with a collection Γ of distinguished covers, we often omit further explicit reference to Γ, and call them *admissible covers*. We modify notions like bi-interpretation etc., to be with respect to Γ. In particular, the notation $\mathrm{Hom}_{\mathcal{T}}$ refers to admissible covers and admissible maps.

Definition 3.1.3. Let $i : \mathcal{T} \to \mathcal{S}$ be an interpretation, and let Γ be a collection of distinguished covers of $\mathcal{T}$. We say that the interpretation i is 2-*stable* (with respect to Γ) if for every two internal covers $\mathcal{T}_1, \mathcal{T}_2$ in Γ over each $\mathcal{T}$-structure A, the natural map $i(\mathrm{Hom}_{\mathcal{T}_A}(\mathcal{T}_1, \mathcal{T}_2)) \to \mathrm{Hom}_{\mathcal{S}_A}(i(\mathcal{T}_1), i(\mathcal{T}_2))$ is an equivalence.

If Γ is omitted, we take it to be all definable groupoids in $\mathcal{T}$, and all definable morphisms among them.

The expression $i(\mathrm{Hom}_{\mathcal{T}}(\mathcal{T}_1, \mathcal{T}_2))$ makes sense, since, as we had noted above, $\mathrm{Hom}_{\mathcal{T}}(\mathcal{T}_1, \mathcal{T}_2)$ is definable in $\mathcal{T}$.

Proposition 3.1.4. *A stable interpretation* $i : \mathcal{T}_1 \to \mathcal{T}_2$ *is* 2-*stable.*

Proof. We may replace $\mathcal{T}$ by $\mathcal{T}_A$, and thus assume that $A = \varnothing$. Let $\mathcal{T}_1, \mathcal{T}_2$ be internal covers of $\mathcal{T}$. The statement is invariant when replacing each cover with a bi-interpretable one (over $\mathcal{T}$). Hence, we may assume that $\mathcal{T}_1 = \mathcal{T}_{\boldsymbol{G}_1}$ and $\mathcal{T}_2 = \mathcal{T}_{\boldsymbol{G}_2}$, the covers associated to definable connected groupoids $\boldsymbol{G}_1, \boldsymbol{G}_2$ in $\mathcal{T}$.

According to Proposition 2.3.1, we may choose $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ so that a stable interpretation of $i(\mathcal{T}_1)$ in $i(\mathcal{T}_2)$ corresponds to a definable map of groupoids from $i(\boldsymbol{G}_2)$ to $i(\boldsymbol{G}_1)$. Since i is stable, this map comes from a map in $\mathcal{T}$ (and similarly for morphisms). $\square$

The definition of a 2-cover is analogous to that of an internal cover, as formulated in Proposition 2.1.2:

Definition 3.1.5. A 2-*internal cover* of a theory $\mathcal{T}$ consists of a theory $\mathcal{T}^*$, a collection Γ of families of internal covers of $\mathcal{T}^*$ over $\mathcal{T}$, and a stable embedding $\mathcal{T} \to \mathcal{T}^*$ such that $\langle \mathcal{T}^*, \Gamma \rangle$ admits a 2-stable interpretation p in $\mathcal{T}_A$, over $\mathcal{T}$ (for some set of parameters A).

More explicitly, we require that each internal cover in Γ is bi-interpretable, over parameters in $\mathcal{T}^*$, with one coming from $\mathcal{T}$, in a manner coherent with interpretations over $\mathcal{T}^*$. Or, via the equivalence with groupoids, that for each family of definable groupoids in Γ there is a set of parameters B in $\mathcal{T}^*$ such that each groupoid in the family is equivalent, over B, to one coming from $\mathcal{T}$ (again, in a coherent manner).

As in the 1-dimensional case, the typical examples come from higher-dimensional groupoids, which we review next.

3.2. *Higher categories and higher groupoids.* We recall a few definitions from homotopy theory and higher category theory, adapted to our language and setup. Though our main references are [8; 9], the ideas seem to originate in [5] (and in the main case of groupoids, much more classically). We are interested in the notions of n-category and n-groupoid discussed in [8, §2.3.4] (through most other parts of the paper we are interested in the case $n = 2$, but here it is convenient and harmless to work in general). There, they are defined as special cases of quasicategories and

Kan simplicial sets, respectively, but for us it is more convenient to use terminology that makes explicit the finite nature of these structures. The following definitions are a variant of the description in [8, §2.3.4.9], which gives an equivalent condition (in the case of simplicial sets).

For each $i \in \mathbb{N}$, we denote by $[i]$ the ordered set $\{0, \dots, i\}$. For $k \in [i]$, we identify k with the map $[0] \to [i]$ taking 0 to k (writing k_i if needed), and we let $\hat{k} = \hat{k}_i : [i-1] \to [i]$ be the unique increasing map with k not in the image. We fix a natural number n (one could also allow $n = \infty$ to obtain the usual definitions of quasicategories and spaces, but we do not use them).

Definition 3.2.1. The signature Σ_n of n-simplicial sets consists of

(1) a sort $\boldsymbol{G}_i$, for $0 \le i \le n+1$,

(2) for each weakly increasing map $t : [i] \to [j]$, where $i, j \le n+1$, a function symbol $d_t : \boldsymbol{G}_j \to \boldsymbol{G}_i$.

3.2.2. *Notation.* We define the following auxiliary notation. We let $\boldsymbol{G}_{-1}$ be the one element set. For each $0 \le m \le n+1$, and $i \le m$, the map $d_{\hat{i}} : \boldsymbol{G}_m \to \boldsymbol{G}_{m-1}$ is called the *i-th face map*. We denote by $\partial = \partial_m : \boldsymbol{G}_m \to \boldsymbol{G}_{m-1}^{m+1}$ the Cartesian product of these maps, and by $\partial^{\hat{i}} : \boldsymbol{G}_m \to \boldsymbol{G}_{m-1}^{m}$ the Cartesian product with i omitted. If $g \in \boldsymbol{G}_m$ and $t : [k] \to [m]$, we sometimes write g_t in place of $d_t(g)$ (in particular, for $t = \hat{i}$ or $t = i$).

For $m \ge -1$, the set $\boldsymbol{G}_{m+1}^{\circ}$ of *$(m+1)$-cycles* is the definable subset of ${\boldsymbol{G}_m}^{m+2}$ given by the conjunction of the equations $d_{\hat{i}}(x_j) = d_{\hat{k}}(x_l)$ for all $0 \le j, l \le m+1$, $0 \le i, k \le m$ satisfying $\hat{j} \circ \hat{i} = \hat{l} \circ \hat{k} : [m-1] \to [m+1]$. (So no conditions when $m = 0$. Note that $(m+1)$-cycles are potential boundaries of $(m+1)$-dimensional elements, but are themselves m-dimensional. This is compatible with the notation in [8].)

For each $0 \le i \le m+1$, the set $\Lambda_i^{m+1}(\boldsymbol{G})$ of *i-th $(m+1)$-horns* is the subset of ${\boldsymbol{G}_m}^{m+1}$ defined by the same conditions, with x_i omitted. Hence, the projection $\pi_{\hat{i}} : \boldsymbol{G}_m^{\circ} \to \boldsymbol{G}_{m-1}^{m}$ omitting the i-th coordinate takes values in $\Lambda_i^m(\boldsymbol{G})$.

We extend the notation by inductively setting $\boldsymbol{G}_m = \boldsymbol{G}_m^{\circ}$ for $m > n+1$, and $d_{\hat{i}} : \boldsymbol{G}_m \to \boldsymbol{G}_{m-1}$ the i-th projection $(0 \le i \le m)$. Consequently, all the above notation makes sense for arbitrary natural number m.

Definition 3.2.3. Let $n \ge 1$. The theory $\mathcal{C}_n$ of n-categories in this signature says:

(1) $d_{t \circ s} = d_s \circ d_t$ for s, t composable, d_t is the identity whenever t is. It follows that ∂_m takes values in $\boldsymbol{G}_m^{\circ}$ and $\partial_m^{\hat{i}}$ in $\Lambda_i^m(\boldsymbol{G})$.

(2) For each $0 < m \le n+1$ and each $0 < i < m$, the map $\partial_m^{\hat{i}} : \boldsymbol{G}_m \to \Lambda_i^m(\boldsymbol{G})$ is surjective.

(3) For $m = n+1, n+2$, $\partial_m^{\hat{i}}$ is bijective for each $0 < i < m$.

The theory $\mathcal{G}_n$ of n-groupoids is the extension of $\mathcal{C}_n$ where the conditions above are required to also hold for $i = 0, m$.

Note that by the third condition, the set $\boldsymbol{G}_{n+1}$ is completely determined by the rest of the data. However, it is still convenient to have it for the statement of the axiom. It follows from the axioms that the unique map $d_f : \boldsymbol{G}_0 \to \boldsymbol{G}_m$ is injective, and we use it to identify $\boldsymbol{G}_0$ with its image in each $\boldsymbol{G}_m$, writing a or a^m for $d_f(a)$ (this map assigns to each object a the m-dimensional identity morphisms at a).

The first condition (when $n = \infty$) is the usual definition of a simplicial set, the second is the definition of quasicategory (or space, in the case of a groupoid, where it is called the Kan condition), and the third specifies that the object is an n-category, rather than a quasicategory. By a *definable n-category* or a *definable n-groupoid* in $\mathcal{T}$ we mean an interpretation in $\mathcal{T}$ of the respective theory.

The intuition is, roughly speaking, that the horns represent configurations of (higher) composable arrows, but the composition (represented by the element g) need not be uniquely determined, except on the highest dimension. We refer to the first chapters of [8] for further explanations, but explain how the case $n = 1$ of the formalism recovers usual categories and groupoids:

Example 3.2.4. A category can be viewed as a 1-category in the above sense by taking $\boldsymbol{G}_0$ the set of objects, $\boldsymbol{G}_1$ the set of morphisms, and $\boldsymbol{G}_2$ the set of pairs of composable morphisms (as we are forced by the axioms). The maps $d_{\hat{0}}, d_{\hat{1}} : \boldsymbol{G}_1 \to \boldsymbol{G}_0$ are the codomain and domain maps, the unique map $\boldsymbol{G}_0 \to \boldsymbol{G}_1$ assigns to each object its identity, and the maps $d_{\hat{0}}, d_{\hat{2}}, d_{\hat{1}} : \boldsymbol{G}_2 \to \boldsymbol{G}_1$ are the two projections and the composition. The only nontrivial instances of the third conditions are when $m = 2$ and $i = 1$, which asserts that any two composable arrows have a unique composition, and when $m = 3$, which corresponds to associativity of the composition.

Conversely, each 1-category determines a category by reversing this process (and likewise for groupoids). □

As in the 1-dimensional case, the axioms imply that for $0 < i < n+1$, the relation $\boldsymbol{G}^\circ_{n+1}$ is the graph of a "composition" function $c_i : \Lambda^n_i(\boldsymbol{G}) \to \boldsymbol{G}_n$, by projecting to the i-coordinate. For n-groupoids, we also have such maps for $i = 0, n+1$.

Remark 3.2.5. If $\boldsymbol{G}$ is an n-category, and $m > n$, our extension of the notation determines a canonical way of viewing $\boldsymbol{G}$ as a Σ_m structure, and as such it is an m-category. Consequently, we view $\boldsymbol{G}$ as an m-category for each $m > n$. If $\boldsymbol{G}$ were an n-groupoid, it would similarly be an m-groupoid for $m > n$. □

3.2.6. *Homotopy sets.* The definition of homotopy sets admits a definable version. Let $\boldsymbol{G}$ be an n-groupoid, and let $b \in \boldsymbol{G}^\circ_m$ $(m \geq 0)$. We let $\boldsymbol{S}(\boldsymbol{G}, b) = \partial_m^{-1}(b)$ be the set of elements of $\boldsymbol{G}_m$ with boundary b. For $\alpha, \beta \in \boldsymbol{S}(\boldsymbol{G}, b)$, we write $\alpha \sim \beta$

(or $\alpha \sim_b \beta$) if some $h \in \boldsymbol{G}_{m+1}$ satisfies $h_{\hat{0}} = \alpha$, $h_{\hat{1}} = \beta$ and $h_{\hat{i}} = d_t(\alpha)_{\hat{i}}$, where $t : [m+1] \to [m]$ is the surjective map with $t(1) = 0$ (so h is a homotopy from α to β, relative to the boundary b). This is an equivalence relation by the Kan condition. Note that when $m \geq n$, this relation coincides with equality.

For $a \in \boldsymbol{G}_0$ and $k \geq 0$, we write $\boldsymbol{S}_k(\boldsymbol{G}, a)$ for $\boldsymbol{S}(\boldsymbol{G}, b)$, where b is the constant boundary with value a in $\boldsymbol{G}_k^\circ$. These are a-definable sets, whose elements correspond to the set of pointed maps from the k-sphere to $\boldsymbol{G}$ with base point a (note that $\boldsymbol{K}_0(\boldsymbol{G}, a) = \boldsymbol{G}_0$ does not actually depend on a). The k-th *homotopy set* of $\boldsymbol{G}$ at a is the quotient $\pi_k(\boldsymbol{G}, a) = \boldsymbol{S}_k(\boldsymbol{G}, a)/\sim$ (in the case of usual simplicial sets, this is one of the equivalent definitions by [9, 00W1]).

If $f : \boldsymbol{G} \to \boldsymbol{H}$ is a groupoid map (between definable n-groupoids in the theory $\mathcal{T}$), it commutes with all the structure above, and therefore induces definable maps of sets $\pi_k(f, a) : \pi_k(\boldsymbol{G}, a) \to \pi_k(\boldsymbol{H}, f(a))$.

Definition 3.2.7. A definable map $f : \boldsymbol{G} \to \boldsymbol{H}$ of n-groupoids is a *weak equivalence* if $\pi_k(f, a) : \pi_k(\boldsymbol{G}, a) \to \pi_k(\boldsymbol{H}, f(a))$ is a bijection for all $0 \leq k \leq n$ and $a \in \boldsymbol{G}_0$.

Remark 3.2.8. More explicitly, for nonempty $\boldsymbol{G}$, the map $f : \boldsymbol{G} \to \boldsymbol{H}$ is a weak equivalence if and only if the following conditions are satisfied for each $n \geq k \geq 0$ and each $a \in \boldsymbol{G}_0$:

(1) For every $g_0, g_1 \in \boldsymbol{S}_k(\boldsymbol{G}, a)$, if $f(g_1) \sim f(g_2)$ then $g_1 \sim g_2$.

(2) For every $h \in \boldsymbol{S}_k(\boldsymbol{H}, f(a))$, there is $g \in \boldsymbol{S}_k(\boldsymbol{G}, a)$ with $f(g) \sim h$.

Alternatively, f is a weak equivalence if and only if it induces a surjective map on $\boldsymbol{S}$-classes, i.e., for each $\boldsymbol{G}$-cycle b, and each $v \in \boldsymbol{S}(\boldsymbol{H}, f(b))$, there is $u \in \boldsymbol{S}(\boldsymbol{G}, b)$ with $f(u) \sim v$. (To prove these equivalences, it suffices to show that they hold in each model, where each of these conditions is equivalent to homotopy equivalence [9, 00WV].) □

Definition 3.2.9. The n-groupoids $\boldsymbol{G}_1$ and $\boldsymbol{G}_2$ are *equivalent* if there are weak equivalences $f_1 : \boldsymbol{G}_1 \to \boldsymbol{H}$ and $f_2 : \boldsymbol{G}_2 \to \boldsymbol{H}$ for some $\boldsymbol{H}$.

Example 3.2.10. Let $\boldsymbol{G}$, $\boldsymbol{H}$ be definable groupoids, viewed as definable 1-groupoids as in Example 3.2.4. A map $f : \boldsymbol{G} \to \boldsymbol{H}$ is a functor. For $k = 0$, the first condition in Remark 3.2.8 says that if a, b are objects of $\boldsymbol{G}$, and there is a morphism between $f(a)$ and $f(b)$ in $\boldsymbol{H}$, then there is a morphism from a to b in $\boldsymbol{G}$. The second condition says that every object h of $\boldsymbol{H}$ has a morphism to an object in the image of f. Together, this part implies that f induces a bijection on isomorphism classes.

For $k = 1$, the first condition says that if g_1, g_2 are automorphisms of a such that $f(g_1) = f(g_2)$, then $g_1 = g_2$, i.e., that f is faithful. The second condition says that f is full. Hence, f is a weak equivalence if and only if it is a weak equivalence in the sense of Section 2.2. In particular, our notion of equivalence coincides with that in [4, §3]. □

As in 2.3.2, equivalence of n-groupoids is an equivalence relation: if $\boldsymbol{H}$ and $\boldsymbol{H}'$ witness that $\boldsymbol{G}_2$ is equivalent to $\boldsymbol{G}_1$ and $\boldsymbol{G}_3$, respectively, the pushout $\boldsymbol{H} \otimes_{\boldsymbol{G}_2} \boldsymbol{H}'$ witnesses the equivalence of $\boldsymbol{G}_1, \boldsymbol{G}_3$.

Remark 3.2.11. The group operation on $\pi_k(\boldsymbol{G}, a)$ (for $k > 0$) is also definable, but we will not use this. □

Remark 3.2.12. The equivalence of our definitions of homotopy groups and weak equivalence with other formulations that appear, for example, in [9, 00V2] does not hold in the definable setting, in general. For example, the analogue of Whitehead's theorem [9, 00WV] is usually false (as seen already in the one-dimensional setting). □

3.2.13. *Morphism groupoids.* Our next goal is to define the space of morphisms between two objects a, b of an n-category $\boldsymbol{G}$, and obtain a (weak) version of the Yoneda embedding that makes sense in the definable setting.

Let $\boldsymbol{G}$ be an n-category, and let $a, b \in \boldsymbol{G}_0$ be two objects. As in [8, §1.2.2], we define the Σ_{n-1}-structure $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, b)$ by

$$\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, b)_k = \{g \in \boldsymbol{G}_{k+1} \mid g_0 = a,\, g_{\hat{0}} = b^k\} \quad \text{for } k \leq n. \tag{1}$$

The structure maps are given by $t \mapsto d^{\boldsymbol{G}}{}_{t^+}$, where $t^+ : [u+1] \to [k+1]$ is given by $t^+(i+1) = t(i)+1$ for $i \in [u]$ and $t^+(0) = 0$. It is clear that $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, b)$ is uniformly definable over a, b when $\boldsymbol{G}$ is definable. It follows from [8, §§4.2.1.8, 2.3.4.18, 2.3.4.19] that this structure is equivalent to an $(n-1)$-groupoid, but since we are not working up to equivalence, we need to prove that it is already an $(n-1)$-groupoid by itself (which we do in Proposition 3.2.15 below).

If we fix a "generic" object $v \in \boldsymbol{G}_0$, the assignment $b \mapsto \underline{\mathbf{Hom}}^L(v, b)$ looks like the object part of the Yoneda embedding for usual categories. One could hope that this is part of a higher Yoneda embedding in our situation as well. However, since there is no composition function for morphisms in $\boldsymbol{G}$, such an embedding does not exist as a functor (it exists noncanonically for set-theoretic quasicategories, but not definably). Instead, we have the following situation (explained in [8, §2.1]): There is an n-category $\boldsymbol{G}_{v/}$ [8, §2.3.4.10], defined by $(\boldsymbol{G}_{v/})_k = \{g \in \boldsymbol{G}_{k+1} \mid g_0 = v\}$, and a map of n-categories $\pi : \boldsymbol{G}_{v/} \to \boldsymbol{G}$, given by $\pi(g) = g_{\hat{0}}$. By definition, the fibre of this map over $b \in \boldsymbol{G}_0$ is $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, b)$. Moreover, this map is a *left fibration* [8, §2.1.22]: given $g \in \Lambda^k_i(\boldsymbol{G}_{v/})$, for $i < k$, any "filling" $h \in \boldsymbol{G}_k$ of $\pi(g)$ (so that $\partial^{\hat{i}}(h) = \pi(g)$) can be lifted to a filling $\tilde{h} \in \boldsymbol{G}_{v/}$ with $\partial^{\hat{i}}(\tilde{h}) = g$ and $\pi(\tilde{h}) = h$. It follows from this that the association $b \mapsto \pi^{-1}(b)$ behaves like a functor of b, but this is only precisely true in the homotopy category.

We show that in the case that h above is invertible, the lifting property above holds for our definable version of equivalence. To do this, we show that the map π

behaves like a local system: the fibres can be continued along (suitable) contractible pieces. The pieces we have in mind are defined as follows:

Definition 3.2.14. The simplicial set D^l, for $l \geq 0$, is defined by $D^l_k = \{0, \ldots, l\}^{\{0,\ldots,k\}}$ (all maps, not necessarily increasing, from $[k]$ to $[l]$), with structure maps given by composition.

We often write elements of $D^l{}_k$ as words of length $k+1$ in the "digits" $0, \ldots, l$. By the usual Yoneda lemma, maps $D^l \to D^m$ correspond (via composition) to functions $\{0, \ldots, l\} \to \{0, \ldots, m\}$. Note that homotopically, all these maps are weak equivalences, and in particular the map to the point D^0, so that all D^l are contractible.

We now extend the definition of morphisms. For $\boldsymbol{G}$ a definable n-category, let $a \in \boldsymbol{G}_0$ be an object, and let $f : D^l \to \boldsymbol{G}$ be a map of simplicial sets (perhaps over parameters). We define a Σ_{n-1}-structure $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$ as follows: for each $k \leq n$,

$$\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)_k = \{\langle g, e\rangle \in \boldsymbol{G}_{k+1} \times D^l{}_k \mid g_0 = a, g_{\hat{0}} = f(e)\} \tag{2}$$

with structure maps given as before by $\langle g, e\rangle \mapsto \langle d_{t^+}(g), e \circ t\rangle$ for each weakly increasing function $t : [u] \to [k]$. In other words, $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$ is the pullback under f of the map $\pi : \boldsymbol{G}_{v/} \to \boldsymbol{G}$ described above. For $l = 0$ and f mapping the point D^0 to b, we recover the previous definition. In general, the projection determines a map $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f) \to D^l$ of simplicial sets, which can be viewed as the "restriction" of π to D^l. If $h : D^r \to D^l$ is a map of simplicial sets, there is an induced map $\hat{h} : \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f \circ h) \to \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$, given by $\hat{h}(\langle g, e\rangle) = \langle g, h(e)\rangle$.

Proposition 3.2.15. *Let $\boldsymbol{G}$, $a \in \boldsymbol{G}_0$ and $f : D^l \to \boldsymbol{G}$ be as above.*

(1) *The structure $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$ is an $(n-1)$-groupoid.*

(2) *For each map $h : \{0, \ldots, r\} \to \{0, \ldots, l\}$ (identified with the corresponding map $D^r \to D^l$), the induced map $\hat{h} : \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f \circ h) \to \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$ is a weak equivalence.*

Proof. (1) Let $\boldsymbol{H} = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$. It is clear that $\boldsymbol{H}$ is a simplicial definable set. To check the Kan condition, we prove a stronger claim, namely, that the projection $\pi : \boldsymbol{H} \to D^l$ is a Kan fibration: given a horn element $h \in \Lambda^m_i(\boldsymbol{H})$ and an element $d \in D^l{}_m$ with $\partial^{\hat{i}}_m(d) = \pi(h)$, there is $\tilde{d} \in \boldsymbol{H}_m$ with $\pi(\tilde{d}) = d$ and $\partial^{\hat{i}}_m(\tilde{d}) = h$.

Let $h \in \Lambda^m_i(\boldsymbol{H})$ be a horn element as above, with $0 \leq i \leq m \leq n+2$. Such an element is given by a matching sequence of elements $h^j = \langle g^j, e^j\rangle$, for $j \in [m]$, $j \neq i$, with $g^j \in \boldsymbol{G}_m$ and with $\pi(h) = \langle e^0, \ldots, e^m\rangle$ an element of $\Lambda^m_i(D^l)$. In D^l, each such horn element comes from a *unique* element of $D^l{}_m$. Let $e \in D^l{}_m$ be this element, and let $g^{-1} = f(e) \in \boldsymbol{G}_m$.

We claim that $\tilde{g} = \langle g^{-1}, g^0, \dots, g^m\rangle \in \boldsymbol{G}_m^{m+1}$ is in $\Lambda_{i+1}^{m+1}(\boldsymbol{G})$. To show that, we need to show that if $\hat{b} \circ \hat{a} = \hat{d} \circ \hat{c} : [m-1] \to [m+1]$ for some $b, d \in [m+1]$, $b, d \neq i+1$ and $a, c \in [m]$, then $g^{b-1}{}_{\hat{a}} = g^{d-1}{}_{\hat{c}}$.

Assume first that $a, b, c, d \geq 1$. The assumption on h implies that

$$h^{b-1}{}_{\widehat{a-1}_{m-1}} = h^{d-1}{}_{\widehat{c-1}_{m-1}} \tag{3}$$

whenever a, b, c, d satisfy

$$\widehat{b-1}_m \circ \widehat{a-1}_{m-1} = \widehat{d-1}_m \circ \widehat{c-1}_{m-1}. \tag{4}$$

Equation (3) implies that

$$g^{b-1}{}_{\widehat{a-1}^{+}_{m-1}} = g^{d-1}{}_{\widehat{c-1}^{+}_{m-1}}$$

under this condition. But $\hat{j}^{+}_{m-1} = \widehat{j+1}_m$ for all $j \in [m-1]$, so we find that $g^{b-1}{}_{\hat{a}_m} = g^{d-1}{}_{\hat{c}_m}$ whenever equation (4) holds. But equation (4) is equivalent to

$$\hat{b}_{m+1} \circ \hat{a}_m = \hat{d}_{m+1} \circ \hat{c}_m \tag{5}$$

so we obtain the required condition when $a, b, c, d \geq 1$.

If $b = 0$ or $d = 0$, the corresponding element of $\boldsymbol{G}_m$ is g^{-1}. In this case, the condition follows from the definition of $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$. For example, if $b = 0$ we must have $c = 0$ and $d = a+1$, so we need to show that $g^a{}_{\hat{0}} = f(e)_{\hat{a}} = f(e_{\hat{a}}) = f(e^a)$, and we are done. If $a = 0$ or $c = 0$, the condition forces $b = 0$ or $d = 0$, so we are back to the same case.

This concludes the proof that $\tilde{g} \in \Lambda^{m+1}_{i+1}(\boldsymbol{G})$. If $i < m$, the Kan condition on $\boldsymbol{G}$ implies that we may find $g \in \boldsymbol{G}_{m+1}$ restricting to $\tilde{g}$. It follows that $g_0 = a$ and $g_{\hat{0}} = f(e)$, so that $\langle g, e\rangle$ solves the lifting problem. It follows from [8, §1.2.5.1] that the case $i < m$ is sufficient.

When $m = n$ or $m = n+1$, the injectivity follows similarly from injectivity for $\boldsymbol{G}$ (and for D^l).

(2) We use Remark 3.2.8. An element in $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f \circ h)_0$ is given by $g \in \boldsymbol{G}_1$ with $g_0 = a$ and $g_1 = f(h(e))$, where $e \in [u]$. Assume that $\langle s, c\rangle, \langle t, d\rangle \in \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f \circ h)_k$ satisfy $s_{\hat{0}} = t_{\hat{0}} = f(h(c)) = f(h(d)) = h(e)$ and $s_{\hat{i}} = t_{\hat{i}} = g$ for $k \geq i > 0$, so that they are elements of $\boldsymbol{S}_k(\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f \circ h))$. Assume also that we are given some $w \in \boldsymbol{G}_{k+2}$ satisfying $w_{\hat{1}} = s$, $w_{\hat{2}} = t$ and $w_{\hat{i}} = g$ for $i > 2$, and some $v \in D^u{}_{k+1}$ with $v_{\hat{0}} = c$, $v_{\hat{1}} = d$ and $v_{\hat{i}} = e$ for $i > 1$, and with $f(h(v)) = w_{\hat{0}}$ (this is a homotopy from $\langle s, c\rangle$ to $\langle t, d\rangle$). Then $\langle w, h(v)\rangle$ is a homotopy from $\langle s, h(c)\rangle$ to $\langle t, h(d)\rangle$. The argument for $k = 0$ is similar (using that D^u is connected).

For the second condition of Remark 3.2.8, let $g \in \boldsymbol{G}_1$ be such that $g_0 = a$, $g_1 = f(e)$ for some $e \in [l]$, so that $b = \langle g, e\rangle$ represents a basepoint of $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$, and let $\langle s, c\rangle \in \boldsymbol{S}_k(\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f), b)$. Then $c \in \boldsymbol{S}_k(D^l, e)$ is the constant function e.

Let $e' \in [u]$, and let $\gamma \in D^l$ be some path from $h(e')$ to e. By the Kan condition above, there is an element s' of $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, f)$ above γ, restricting to s. This s' serves as a homotopy from s to an element over $h(e')$, which is thus in the image of $\hat{h}$. $\square$

Corollary 3.2.16. *Let $\boldsymbol{G}$ be an n-category, $v, a, b \in \boldsymbol{G}_0$ objects. Each isomorphism $t \in \boldsymbol{G}_1$ from a to b determines an equivalence $e_t : \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, a) \to \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, b)$. If $s \in \boldsymbol{G}_1$ is another isomorphism from a to b, each isomorphism $m : t \to s$ determines an isomorphism $e_m : e_t \to e_s$.*

Proof. Apply Proposition 3.2.15 to maps from D^1 and from D^2 determined by t, s and e. $\square$

We describe the equivalence explicitly in the 2-dimensional case, which is most relevant for us:

Example 3.2.17. Let $\boldsymbol{G}$ be a 2-groupoid, $v, a, b \in \boldsymbol{G}_0$ and $f \in \boldsymbol{G}_1$ with $f_0 = a$ and $f_1 = b$. Since $\boldsymbol{G}$ is a groupoid, there is $h \in \boldsymbol{G}_2$ (not necessarily unique) with $h_{\hat{0}} = f$ and $h_{\hat{1}} = b$. We denote $f^{-1} = h_{\hat{2}}$. By the 2-groupoid axioms, h has a uniquely determined inverse h^{-1}. Let $\gamma : D^1 \to \boldsymbol{G}$ be the unique map with $\gamma(101) = h$, so that $\gamma(01) = g$ and $\gamma(10) = g^{-1}$. Then $\boldsymbol{H} = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, \gamma)$ can be described as follows:

(1) $\boldsymbol{H}_0 = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, a)_0 \amalg \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, b)_0$ (this is just the union if $a \neq b$, but if $a = b$ we take disjoint copies).

(2) Let $\boldsymbol{X} = \{g \in \boldsymbol{G}_2 \mid g_0 = v, g_{\hat{0}} = f\}$, and let $\boldsymbol{X}^{-1} = \{g \in \boldsymbol{G}_2 \mid g_0 = v, g_{\hat{0}} = f^{-1}\}$ (again taking disjoint copies if $f = f^{-1}$). Then

$$\boldsymbol{H}_1 = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, a)_1 \amalg \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, b)_1 \cup \boldsymbol{X} \cup \boldsymbol{X}^{-1}$$

with $d^{\boldsymbol{X}}_{\hat{0}} = d^{\boldsymbol{G}}_{\hat{1}}$, $d^{\boldsymbol{X}}_{\hat{1}} = d^{\boldsymbol{G}}_{\hat{2}}$ and vice versa for $\boldsymbol{X}^{-1}$ (and the structure coming from $\underline{\mathbf{Hom}}^L$ on the other parts).

(3) Composition is defined again as in $\underline{\mathbf{Hom}}^L$ on the corresponding parts. The composition of $h \circ g$ for $g \in \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, a)$ and $h \in \boldsymbol{X}$ is the composition in $\boldsymbol{G}$ of the three elements $g, h, i \in \boldsymbol{G}_2$, where i is the identity morphism of the object f of $\underline{\mathbf{Hom}}^R_{\boldsymbol{G}}(a, b)$. Similarly for the compositions $g' \circ h$, $h' \circ g'$, $g \circ h'$, $h \circ h'$ and $h' \circ h$, for $g' \in \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(v, b)$ and $h' \in \boldsymbol{X}^{-1}$ (in each case, the two elements of $\boldsymbol{G}_2$ along with i form three faces of a 2-horn, with vertices a, a, b, v or a, b, b, v, and the result is the uniquely determined fourth face).

It is clear, by construction, that each of the inclusions of $\underline{\mathbf{Hom}}^L(v, a)$ and of $\underline{\mathbf{Hom}}^L(v, b)$ into $\boldsymbol{H}$ determine fully faithful functors. As in the general proof, they are also essentially surjective by the Kan property. $\square$

Corollary 3.2.18. *Let $\boldsymbol{G}$ be a 2-groupoid, $\gamma : D^2 \to \boldsymbol{G}$ a fixed map, and $a \in \boldsymbol{G}_0$ a fixed vertex. Then*

$$\underline{\mathbf{Hom}}^L(a, \gamma) = \underline{\mathbf{Hom}}^L(a, \gamma \circ \hat{2}) \otimes_{\underline{\mathbf{Hom}}^L(a, \gamma\circ 1)} \underline{\mathbf{Hom}}^L(a, \gamma \circ \hat{0})$$

(*canonical isomorphism*), *and*

$$\underline{\mathbf{Hom}}^L(a, \gamma \circ \hat{1}) = \underline{\mathbf{Hom}}^L(a, \gamma \circ \hat{0}) \circ_{\underline{\mathbf{Hom}}^L(a, \gamma\circ 1)} \underline{\mathbf{Hom}}^L(a, \gamma \circ \hat{2}).$$

In other words, the composition of two morphism groupoids (in the sense of 2.3.2) is given by composition in the homotopy category.

Proof. By definition, both sides have the same sets of objects. Proposition 2.2.9 provides the required map, and since on both sides we also have a weak equivalence (by the second part of Proposition 2.2.9 and by Proposition 3.2.15), this map is an isomorphism. The second part again follows directly from the definition, as both sides are the restriction to the same set of objects. □

3.3. ***The theory associated with a groupoid.*** We continue to fix $n \in \mathbb{N}$. To each definable n-groupoid in the theory $\mathcal{T}$ we define an associated expansion $\mathcal{T}_{\boldsymbol{G}}$ of $\mathcal{T}$, directly generalising (a variant of) the one-dimensional case (Construction 2.2.4).

Definition 3.3.1. Let $\boldsymbol{G}$ be a definable n-groupoid in a theory $\mathcal{T}$. The expansion $\mathcal{T}_{\boldsymbol{G}}$ of $\mathcal{T}$ is obtained by adding additional sorts $\boldsymbol{G}_i^*$ for $0 \le i \le n+1$, function symbols $e_i : \boldsymbol{G}_i \to \boldsymbol{G}_i^*$, a constant symbol $* \in \boldsymbol{G}_0^*$, and the axioms expressing:

(1) $\boldsymbol{G}^*$ is an n-groupoid, and e_* is a map of simplicial sets (i.e., commutes with the structure maps). We identify $\boldsymbol{G}$ with its image.

(2) $\boldsymbol{G}_0^* = \boldsymbol{G}_0 \cup \{*\}$.

(3) The inclusion of $\boldsymbol{G}$ in $\boldsymbol{G}^*$ is a weak homotopy equivalence (Remark 3.2.8), and an isomorphism onto the full subgroupoid of $\boldsymbol{G}^*$ spanned by $\boldsymbol{G}_0$.

For each natural number r, there is a definable family $\Gamma_r = \underline{\mathbf{Hom}}^L(*, f)$ of groupoids, parametrised by the definable set of maps $f : D^r \to \boldsymbol{G}^*$. This is our collection Γ of admissible groupoids, in the sense of Definition 3.1.3.

We note that our choice of Γ does satisfy the assumption on composition, by Corollary 3.2.18.

As in the one-dimensional case, each object $a \in \boldsymbol{G}_0$ determines an interpretation ω_a over $\mathcal{T}_a$ determined by the requirement that $\omega_a(*) = a$, $\omega_a({\boldsymbol{G}^*}_i) = \boldsymbol{G}_i$ for $i = 1, 2$ and similarly for the face maps. We would like to show that the ω_a are objects in the 2-groupoid associated with $\mathcal{T}_{\boldsymbol{G}}$ over $\mathcal{T}$, namely:

Proposition 3.3.2. *For every object $a \in \boldsymbol{G}_0$, the interpretation $\omega_a : \mathcal{T}_{\boldsymbol{G}} \to \mathcal{T}_a$ is 2-stable. In particular, $\langle \mathcal{T}_{\boldsymbol{G}}, \Gamma \rangle$ is a 2-internal cover of $\mathcal{T}$.*

Proof. We need to show that over some parameter u, each Γ-admissible groupoid $\boldsymbol{H}$ is equivalent to $\omega_a(\boldsymbol{H})$, over some parameters from $\mathcal{T}$. Let u be any element of $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*, a)_0$ (it is consistent that such a u exists: for any model M of $\mathcal{T}$ such that $a \in \boldsymbol{G}_0(\mathrm{M})$, $\mathrm{M} \circ \omega_a$ is a model of $\mathcal{T}_{\boldsymbol{G}}$ for which this set is nonempty).

Let $f : D^r \to \boldsymbol{G}^*$ be a map, and assume first that for some $i \in [r]$, $b = f(i) \in \boldsymbol{G}_0$. By the second item of Proposition 3.2.15, $\boldsymbol{H} = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*, f)$ is equivalent (over no additional parameters) to $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*, b)$, so we may assume that $f = b$. We may also assume that b is in the same connected component as a, because otherwise $\boldsymbol{H}$ is empty. According to Corollary 3.2.16, it follows that $\boldsymbol{H}$ is equivalent to $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*, a)$. Again according to (a dual version of) Corollary 3.2.16, the fixed element u determines an equivalence from $\boldsymbol{H}$ to $\omega_a(\boldsymbol{H}) = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, a)$.

The remaining case is when f is the constant map $*$, so that $\boldsymbol{H} = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*, *)$, and $\omega_a(\boldsymbol{H}) = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, a)$. The same argument as above shows that both are equivalent to $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*, a)$ over u. □

We would like to prove that the association $a \mapsto \omega_a$ is the object part of an assignment that recovers (up to equivalence) $\boldsymbol{G}$. To do that, we need to define the 2-groupoid which is the target of this assignment. This will be the analogue of $I_{\mathcal{T}*/\mathcal{T}}$ from the one-dimensional case (2.2.2).

Definition 3.3.3. Let $\mathcal{T}^*$ be a stable expansion of a theory $\mathcal{T}$, with admissible family of distinguished internal covers Γ. The 2-groupoid associated to this datum is defined as follows:

(1) Objects are 2-stable interpretations of $\langle \mathcal{T}^*, \Gamma \rangle$ in $\mathcal{T}$, over $\mathcal{T}$.

(2) If x, y are two objects as above, a morphism $u : x \to y$ is given by a bi-interpretation $u_{\mathcal{T}'} : x(\mathcal{T}') \to y(\mathcal{T}')$ over $\mathcal{T}$, for each admissible internal cover $\mathcal{T}' \in \Gamma$. These bi-interpretations are given with isomorphisms $c_i : u_{\mathcal{T}_2} \circ x(i) \to y(i) \circ u_{\mathcal{T}_1}$ for every admissible interpretation $i : \mathcal{T}_1 \to \mathcal{T}_2$ between admissible covers $\mathcal{T}_1, \mathcal{T}_2$ in Γ (uniformly in families).

The isomorphisms are required to satisfy $c_{j \circ i} = y(j)(c_i) \circ c_j(x(i))$ for admissible interpretations $i : \mathcal{T}_1 \to \mathcal{T}_2$, $j : \mathcal{T}_2 \to \mathcal{T}_3$ as above (these make sense since, by definition, the c_i are definable maps in $y(\mathcal{T}_2)$, $y(j)$ is an interpretation of $y(\mathcal{T}_2)$ in $y(\mathcal{T}_3)$ and c_j is a map between interpretations of $x(\mathcal{T}_2)$, so can be applied to definable sets of the form $x(i)$).

(3) The 2-morphisms with edges $u : x \to y$, $v : y \to z$ and $w : x \to z$ are given by isomorphisms $v \circ u \to w$, all over $\mathcal{T}$.

(4) The "2-composition" of the 2-morphisms $\alpha : v \circ u \to w$, $\beta : s \circ v \to r$ and $\gamma : r \circ u \to t$ is given by

$$s \circ w \xrightarrow{s \cdot \alpha^{-1}} s \circ (v \circ u) = (s \circ v) \circ u \xrightarrow{\beta \cdot u} r \circ u \xrightarrow{\gamma} t,$$

where $\cdot$ stands for pointwise application (or *horizontal composition*) as above.

Applying the definition with $\mathcal{T}$ replaced by $\mathcal{T}_A$, for a $\mathcal{T}$-structure A, we obtain a 2-groupoid for each such structure A, which we denote $I^2(A) = I^2{}_{\mathcal{T}^*/\mathcal{T}}(A)$.

Using the equivalence between internal covers and definable groupoids, this can be described in terms of definable groupoids. We give an explicit description in 3.3.6 below.

Remark 3.3.4. Let $\mathcal{C}$ be the category of definable groupoids in $\mathcal{T}$, with weak equivalences as morphisms. We may form its bicategory of *cospans* for this category, as in [9, 0084]. By Proposition 2.3.4, it is equivalent (as a bicategory) to the category of internal covers and bi-interpretations. The 2-category in Definition 3.3.3 can be viewed as the *Duskin nerve* [9, 009T] of this bicategory (clear from the description in [9, 00A1]). In particular, it follows that this is indeed a 2-category. $\square$

Proposition 3.3.5. *Let $\boldsymbol{G}$ be a definable 2-groupoid in a theory $\mathcal{T}$, and $\mathcal{T}_{\boldsymbol{G}} = \langle \mathcal{T}_{\boldsymbol{G}}, \Gamma \rangle$ the corresponding 2-internal cover. The association $a \mapsto \omega_a$ extends to a map $\omega : \boldsymbol{G}(A) \to I^2{}_{\mathcal{T}_G/\mathcal{T}}(A)$ of 2-groupoids, compatible with extensions of the structure A.*

Proof. Proposition 3.3.2 shows that for all $a \in \boldsymbol{G}_0(A)$, ω_a is indeed an object of $I^2(A)$. Given $t : a \to b$ in $\boldsymbol{G}_1(A)$, we define $\omega_t : \omega_a \to \omega_b$ as follows.

Let $\boldsymbol{H}$ be an admissible groupoid. As in the proof of Proposition 3.3.2, we assume that $\boldsymbol{H} = \boldsymbol{H}_c = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*, c)$ for some $c \in \boldsymbol{G}_0(A)$, so that $\omega_a(\boldsymbol{H}) = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(a, c)$ and $\omega_b(\boldsymbol{H}) = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(b, c)$. By Corollary 3.2.16, t induces an (admissible) equivalence from $\omega_a(\boldsymbol{H})$ to $\omega_b(\boldsymbol{H})$, which we take to be $\omega_t(\boldsymbol{H})$. Our definition (and construction) ensures the compatibility under admissible maps $\boldsymbol{H} \to \boldsymbol{H}'$.

Similarly, let $\alpha \in \boldsymbol{G}_2(A)$, with edges $r : a \to b$, $s : b \to c$ and $t : a \to c$. We need to construct an isomorphism (over $\mathcal{T}$) from $\omega_s \circ \omega_r$ to ω_t. Consider the map $f : D^2 \to \boldsymbol{G}$ determined by α. We have $f(01) = r$, $f(12) = s$ and $f(02) = t$, so that for each object $d \in \boldsymbol{G}_0(A)$, the equivalence $\omega_r(\boldsymbol{H}_d) : \omega_a(\boldsymbol{H}_d) \to \omega_b(\boldsymbol{H}_d)$ is given by $\underline{\mathbf{Hom}}^L(f \circ h_{01}, d)$, where $h_{01} : [1] \to [2]$ is the inclusion (and similarly for s, t). Hence, $\underline{\mathbf{Hom}}^L(f, d)$ represents the composition $\omega_s \circ \omega_r$, and restriction to ω_t provides the required map.

This completes the construction of ω. The proof that this is a map of 2-groupoids (i.e., that it commutes with composition) is similar to the above, using D^4 in place of D^3, and the fact that it commutes with extension of scalars is obvious. $\square$

3.3.6. Our main goal is to prove that the map ω constructed in Proposition 3.3.5 is a weak equivalence. Similarly to the 1-dimensional case, it is generally only true in a model. As a preparation, we consider more explicitly the structure of I^2 from Definition 3.3.3, from a definable groupoid point of view.

Let ω_1 and ω_2 be two objects of $I^2_{\mathcal{T}^*/\mathcal{T}}$, i.e., 2-stable interpretations of $\mathcal{T}^*$ in $\mathcal{T}$. An isomorphism from ω_1 to ω_2 over $\mathcal{T}$ is given, according to Proposition 2.3.1, by a family $\boldsymbol{K}(\boldsymbol{H})$ of groupoids in $\mathcal{T}$, for each admissible groupoid $\boldsymbol{H}$ in $\mathcal{T}^*$, along with weak equivalences $u_i(\boldsymbol{H}) : \omega_i(\boldsymbol{H}) \overset{\sim}{\dashrightarrow} \boldsymbol{K}(\boldsymbol{H})$, all definable uniformly in $\boldsymbol{H}$. Given another admissible groupoid $\boldsymbol{H}'$, an admissible interpretation from $\mathcal{T}^*{}_{\boldsymbol{H}'}$ to $\mathcal{T}^*{}_{\boldsymbol{H}}$ is given, again by Proposition 2.3.1, by an admissible groupoid $\boldsymbol{X}$ and admissible maps $f : \boldsymbol{H} \to \boldsymbol{X}$ and $g : \boldsymbol{H}' \overset{\sim}{\dashrightarrow} \boldsymbol{X}$.

According to Definition 3.3.3, we are provided with definable isomorphisms (realising the isomorphisms c_i there, via Proposition 2.3.4)

$$t_{\boldsymbol{X},\boldsymbol{K}} : \boldsymbol{K}(\boldsymbol{H}) \circ_{\omega_1(\boldsymbol{H})} \omega_1(\boldsymbol{X}) \xrightarrow{\sim} \omega_2(\boldsymbol{X}) \circ_{\omega_2(\boldsymbol{H}')} \boldsymbol{K}(\boldsymbol{H}') \tag{6}$$

uniformly definable in $\boldsymbol{X}$, $\boldsymbol{K}$ (and the associated embeddings), and restricting to the identity on $\omega_1(\boldsymbol{H}')$ and on $\omega_2(\boldsymbol{H})$. The situation is depicted in the top part of the following diagram:

$$\begin{tikzcd}
\omega_1(\boldsymbol{H}) \arrow[rr, dashed, "u_1(\boldsymbol{H})", "\sim"'] \arrow[dd, "\omega_1(f)"'] & & \boldsymbol{K}(\boldsymbol{H}) & & \omega_2(\boldsymbol{H}) \arrow[ll, "u_2(\boldsymbol{H})"'] \arrow[dlll, bend right] \arrow[dddl, bend left] \arrow[dd, "\omega_2(f)"] \\
& \boldsymbol{K}(\boldsymbol{H}) \circ_{\omega_1(\boldsymbol{H})} \omega_1(\boldsymbol{X}) \arrow[ddrr, "t_{\boldsymbol{X},\boldsymbol{K}}"', "\sim"] & & & \\
\omega_1(\boldsymbol{X}) & & & & \omega_2(\boldsymbol{X}) \\
& & & \omega_2(\boldsymbol{X}) \circ_{\omega_2(\boldsymbol{H}')} \boldsymbol{K}(\boldsymbol{H}') & \\
\omega_1(\boldsymbol{H}') \arrow[uu, dashed, "\omega_1(g)", "\sim"'] \arrow[uuur, dashed, bend left, "\sim"] \arrow[urrr, dashed, bend left, "\sim"] \arrow[rr, dashed, "\sim", "u_1(\boldsymbol{H}')"'] \arrow[dd] & & \boldsymbol{K}(\boldsymbol{H}') & & \omega_2(\boldsymbol{H}') \arrow[ll, "u_2(\boldsymbol{H}')"'] \arrow[uu, dashed, "\sim", "\omega_2(g)"'] \arrow[dlll, bend right] \arrow[dddl, bend left] \arrow[dd] \\
& \boldsymbol{K}(\boldsymbol{H}') \circ_{\omega_1(\boldsymbol{H}')} \omega_1(\boldsymbol{Y}) \arrow[ddrr, "t_{\boldsymbol{Y},\boldsymbol{K}}"', "\sim"] & & & \\
\omega_1(\boldsymbol{Y}) & & & & \omega_2(\boldsymbol{Y}) \\
& & & \omega_2(\boldsymbol{Y}) \circ_{\omega_2(\boldsymbol{H}'')} \boldsymbol{K}(\boldsymbol{H}'') & \\
\omega_1(\boldsymbol{H}'') \arrow[uu, dashed, "\sim"'] \arrow[uuur, dashed, bend left, "\sim"] \arrow[urrr, dashed, bend left, "\sim"] \arrow[rr, dashed, "\sim", "u_1(\boldsymbol{H}'')"'] & & \boldsymbol{K}(\boldsymbol{H}'') & & \omega_2(\boldsymbol{H}'') \arrow[ll, "u_2(\boldsymbol{H}'')"] \arrow[uu, dashed, "\sim"]
\end{tikzcd}$$

If $\boldsymbol{Y}$ determines a map to $\mathcal{T}^*{}_{\boldsymbol{H}'}$ from $\mathcal{T}^*{}_{\boldsymbol{H}''}$ for a further groupoid $\boldsymbol{H}''$, we have the maps

$$\begin{aligned} t_{\boldsymbol{X},\boldsymbol{K}} \otimes_{\omega_1(\boldsymbol{H}')} \mathbf{1}_{\omega_1(\boldsymbol{Y})} : \boldsymbol{K}(\boldsymbol{H}) \circ_{\omega_1(\boldsymbol{H})} \omega_1(\boldsymbol{X}) \circ_{\omega_1(\boldsymbol{H}')} \omega_1(\boldsymbol{Y}) & \\ \xrightarrow{\sim} \omega_2(\boldsymbol{X}) \circ_{\omega_2(\boldsymbol{H}')} \boldsymbol{K}(\boldsymbol{H}') \circ_{\omega_1(\boldsymbol{H}')} \omega_1(\boldsymbol{Y}) & \end{aligned} \tag{7}$$

and

$$\mathbf{1}_{\omega_2(\boldsymbol{X})} \otimes_{\omega_2(\boldsymbol{H}')} t_{\boldsymbol{Y},\boldsymbol{K}} : \omega_2(\boldsymbol{X}) \circ_{\omega_2(\boldsymbol{H}')} \boldsymbol{K}(\boldsymbol{H}') \circ_{\omega_1(\boldsymbol{H}')} \omega_1(\boldsymbol{Y}) \\ \xrightarrow{\sim} \omega_2(\boldsymbol{X}) \circ_{\omega_2(\boldsymbol{H}')} \omega_2(\boldsymbol{Y}) \circ_{\omega_2(\boldsymbol{H}'')} \boldsymbol{K}(\boldsymbol{H}''). \quad (8)$$

The groupoid $\boldsymbol{X} \circ_{\boldsymbol{H}'} \boldsymbol{Y}$ represents the composition of interpretations, and

$$\omega_i(\boldsymbol{X} \circ_{\boldsymbol{H}'} \boldsymbol{Y}) = \omega_i(\boldsymbol{X}) \circ_{\omega_i(\boldsymbol{H}')} \omega_i(\boldsymbol{Y})$$

(canonical identification), since ω_i is an interpretation. Under this identification, we require that

$$(\mathbf{1}_{\omega_2(\boldsymbol{X})} \otimes_{\omega_2(\boldsymbol{H}')} t_{\boldsymbol{Y},\boldsymbol{K}}) \circ (t_{\boldsymbol{X},\boldsymbol{K}} \otimes_{\omega_1(\boldsymbol{H}')} \mathbf{1}_{\omega_1(\boldsymbol{Y})}) = t_{(\boldsymbol{X} \circ_{\boldsymbol{H}'} \boldsymbol{Y}),\boldsymbol{K}}. \quad (9)$$

Finally, a 2-morphism is determined by a natural isomorphism between two maps as above (one a composition, which we already understand), so it is enough to describe those. Let ω_1, ω_2 and $\boldsymbol{K}$ be as above, and let $\boldsymbol{L}$ represent another morphism. A natural isomorphism is then given by a uniform family of isomorphisms $\alpha_{\boldsymbol{H}} : \boldsymbol{K}(\boldsymbol{H}) \to \boldsymbol{L}(\boldsymbol{H})$ over $\omega_i(\boldsymbol{H})$, which intertwine the maps $t_{\boldsymbol{X},\boldsymbol{K}}$ and $t_{\boldsymbol{X},\boldsymbol{L}}$ whenever $\boldsymbol{X}$ represents an interpretation. The 2-composition of three such suitable maps is described as in Definition 3.3.3, with composition replaced by pushouts as appropriate.

Remark 3.3.7. By definition, internality means that there is a nonempty definable *set* (i.e., a 0-groupoid) of isomorphisms between the internal sorts and sorts of the base theory. Similarly, the structure described above includes the description of a $\mathcal{T}^*$-definable 1-groupoid $\underline{\mathbf{Iso}}_{\mathcal{T}}(\widetilde{\mathcal{T}}^*, \widetilde{\mathcal{T}})$ of weak equivalences between admissible covers $\widetilde{\mathcal{T}}^*$ of $\mathcal{T}^*$ and covers $\widetilde{\mathcal{T}}$ of $\mathcal{T}$ (nonempty for some $\widetilde{\mathcal{T}}$ if $\mathcal{T}^*$ is 2-internal). In terms of groupoids, the families $\boldsymbol{K}$ as in 3.3.6 are the objects, and the morphisms are the natural isomorphisms α. Furthermore, this groupoid itself is admissible. □

Example 3.3.8. Let $\mathcal{T}^* = \mathcal{T}_{\boldsymbol{G}}$ as in Proposition 3.3.5, and let $\omega_1 = \omega_a$ and $\omega_2 = \omega_b$ for some $a, b \in \boldsymbol{G}_0(A)$. Let $f : a \to b$ be a morphism in $\boldsymbol{G}(A)$ (identified with the corresponding map from D^1). Given an admissible groupoid $\boldsymbol{H} = \boldsymbol{H}_d = \underline{\mathbf{Hom}}^L(d, *)$, we let $\boldsymbol{K}_f(d) = \boldsymbol{K}_f(\boldsymbol{H}_d) = \underline{\mathbf{Hom}}^L_{\boldsymbol{G}}(d, f)$, with the canonical maps from $\omega_a(\boldsymbol{H}_d) = \underline{\mathbf{Hom}}^L(d, a)$ and $\omega_b(\boldsymbol{H}_d) = \underline{\mathbf{Hom}}^L(d, b)$ (these are weak equivalences by Proposition 3.2.15).

To give $\boldsymbol{K}$ the structure of an isomorphism from ω_a to ω_b, we need to supply the isomorphisms (6). If $\boldsymbol{H}' = \boldsymbol{H}_c = \underline{\mathbf{Hom}}^L(c, *)$ is another admissible groupoid in $\mathcal{T}_{\boldsymbol{G}}$, an admissible isomorphism from $\boldsymbol{H}_c$ to $\boldsymbol{H}_d$ is given by a groupoid $\boldsymbol{X}_g = \underline{\mathbf{Hom}}^L(g, *)$, with $g : c \to d$ in $\boldsymbol{G}$. Seeing as $\omega_x(\boldsymbol{X}_g) = \underline{\mathbf{Hom}}^L(g, x)$ for all x, such a structure consists of a definable family of maps

$$t_{g,f} : \underline{\mathbf{Hom}}^L(d, f) \circ_{\underline{\mathbf{Hom}}^L(d,a)} \underline{\mathbf{Hom}}^L(g, a) \xrightarrow{\sim} \underline{\mathbf{Hom}}^L(g, b) \circ_{\underline{\mathbf{Hom}}^L(c,b)} \underline{\mathbf{Hom}}^L(c, f).$$

Recall that t is the identity on objects, so we only need to define it on morphisms. Let $\langle u, v\rangle$ represent a morphism of $\underline{\mathbf{Hom}}^L(d, f) \circ_{\underline{\mathbf{Hom}}^L(d,a)} \underline{\mathbf{Hom}}^L(g, a)$. Let $h : c \to a$ be the domain of v. By the Kan property, there is a morphism $w \in \underline{\mathbf{Hom}}^L(c, f)$ whose domain is h. Then u, v, w form an element of $\Lambda^3_3(\boldsymbol{G})$, so composition provides a fourth face $y \in \underline{\mathbf{Hom}}^L(g, b)_1$. We let $t_{g,f}(u \otimes v) = y \otimes w$. If w' is a different choice in place of w, then $w' \circ w^{-1}$ is in $\underline{\mathbf{Hom}}^L(c, b)$, so the result represents the same morphism of $\underline{\mathbf{Hom}}^L(g, b) \circ_{\underline{\mathbf{Hom}}^L(c,b)} \underline{\mathbf{Hom}}^L(c, f)$. It is clear that t is well defined on the class $u \otimes v$, and uniformly definable in g, f.

To prove the identity (9), assume we are given another morphism $g' : c' \to c$, corresponding to an admissible interpretation represented by $\boldsymbol{Y} = \underline{\mathbf{Hom}}^L(g', *)$. Let $v' \in \underline{\mathbf{Hom}}^L(g', a) = \omega_1(\boldsymbol{Y})$. Proceeding with the notation above, we need to determine the image of $y \otimes w \otimes v'$ in

$$\underline{\mathbf{Hom}}^L(g, b) \circ_{\underline{\mathbf{Hom}}^L(c,b)} \underline{\mathbf{Hom}}^L(g', b) \circ_{\underline{\mathbf{Hom}}^L(c',b)} \underline{\mathbf{Hom}}^L(c', f).$$

As above, it is given by $y \otimes y' \otimes w'$, where $y' \in \underline{\mathbf{Hom}}^L(g', b) = \omega_2(\boldsymbol{Y})$ and $w' \in \underline{\mathbf{Hom}}^L(c', f) = \boldsymbol{K}_f(\boldsymbol{H}_{c'})$ represent the other two faces of a partial simplex with faces w and v' (this other simplex can be visualised as attached to the previous one at the face w). On the other hand, $\boldsymbol{X} \circ_{\boldsymbol{H}_c} \boldsymbol{Y}$ was identified (as in Corollary 3.2.18) with $\underline{\mathbf{Hom}}^L(g \circ g', *)$, for any composition $g \circ g'$. After choosing such a composition h, $v \otimes v'$ is identified with an element of $\underline{\mathbf{Hom}}^L(h, a)$ and $y \otimes y'$ with an element of $\underline{\mathbf{Hom}}^L(h, b)$, so that they become two faces of the simplex with vertices a, b, c, d, the other two being u and w', so that $u \otimes w' = t_{h,f}(v \otimes v', y \otimes y')$, as required.

Assume now that we are given a map $\gamma : D^2 \to \boldsymbol{G}$ corresponding to an element $w \in \boldsymbol{G}_2$, with edges $f = \gamma(01)$, $g = \gamma(12)$ and $h = \gamma(02)$. Given an element $u \in \boldsymbol{K}_f(\boldsymbol{H}_c) = \underline{\mathbf{Hom}}^L(c, f)$ and $v \in \boldsymbol{K}_g(\boldsymbol{H}_c) = \underline{\mathbf{Hom}}^L(c, g)$ (for an arbitrary $c \in \boldsymbol{G}_0$), the 2-composition applied to u, v and w provides an element of $\boldsymbol{K}_h(\boldsymbol{H}_c) = \underline{\mathbf{Hom}}^L(c, h)$. This process assembles into a family of isomorphisms

$$\alpha_\gamma : \boldsymbol{K}_g(\boldsymbol{H}_c) \circ \boldsymbol{K}_f(\boldsymbol{H}_c) \to \boldsymbol{K}_h(\boldsymbol{H}_c),$$

definable uniformly in γ and c. This completes the description (and a reformulation of the proof) of the map in Proposition 3.3.5 in terms of definable groupoids. □

We are now ready to prove our main result.

Theorem 3.3.9. *Let $\boldsymbol{G}$ be a 2-groupoid defined in a theory $\mathcal{T}$, and let $\mathcal{T}_{\boldsymbol{G}}$ be the associated theory (and admissible covers), as in Definition 3.3.1. Then $\mathcal{T}_{\boldsymbol{G}}$ is a 2-internal cover of $\mathcal{T}$, and for every model* M *of $\mathcal{T}$, the 2-groupoid of* M*-points $I^2_{\mathcal{T}_{\boldsymbol{G}}/\mathcal{T}}(M)$ is weakly equivalent to $\boldsymbol{G}(M)$.*

Proof. The fact that $\mathcal{T}_{\boldsymbol{G}}$ is a 2-internal cover is Proposition 3.3.2. The map from $\boldsymbol{G}$ to $I^2_{\mathcal{T}_{\boldsymbol{G}}/\mathcal{T}}$ was constructed in Proposition 3.3.5, and described in terms of groupoids in Example 3.3.8. We use this description to show that the map is a weak equivalence, using Remark 3.2.8. We assume that we are working over M, and proceed by considering the possible dimensions $0 \leq k \leq 2$.

$\underline{k=0}$: We need to show that any 2-stable interpretation ω of $\mathcal{T}_{\boldsymbol{G}}$ in $\mathcal{T}$ admits a coherent collection of bi-interpretations as in Definition 3.3.3(2) to some ω_a.

Since ω is an interpretation over $\mathcal{T}$, $\omega(\boldsymbol{G}^*)$ is a definable 2-groupoid in $\mathcal{T}$, containing $\boldsymbol{G}$, with the inclusion a weak equivalence. The proof now proceeds exactly as the proof of Proposition 3.3.2, with $\omega(\boldsymbol{G}^*)$ in place of $\boldsymbol{G}^*$.

$\underline{k=1}$: This is the main case, which can be viewed as a definable version of the Yoneda lemma. Let $a, b \in \boldsymbol{G}_0$, and assume we are given an equivalence from ω_a to ω_b. Hence, for every $c \in \boldsymbol{G}_0$ (some parameters), we are given a groupoid $\boldsymbol{K}(c)$ in $\mathcal{T}$ and weak equivalences $\underline{\mathbf{Hom}}^L(c, a) = \omega_a(\boldsymbol{H}_c) \to \boldsymbol{K}(c)$ and $\underline{\mathbf{Hom}}^L(c, b) = \omega_b(\boldsymbol{H}_c) \to \boldsymbol{K}_c$, uniformly in c, along with structure maps (6)

$$t_{g,\boldsymbol{K}} : \boldsymbol{K}(d) \circ_{\underline{\mathbf{Hom}}^L(d,a)} \underline{\mathbf{Hom}}^L(g, a) \to \underline{\mathbf{Hom}}^L(g, b) \circ_{\underline{\mathbf{Hom}}^L(c,b)} \boldsymbol{K}(c)$$

(all notation as in Example 3.3.8, except $\boldsymbol{K}$ is no longer known to be of the given form). We identify $\omega_a(\boldsymbol{H}_c)$, $\omega_b(\boldsymbol{H}_c)$ with their images in $\boldsymbol{K}(c)$.

In particular, we have the identity morphism $\mathbf{1}_a$ of a as an object $\mathbf{1}_a \in \omega_a(\boldsymbol{H}_a)$, and by weak equivalence, an object $f : a \to b$ in $\omega_b(\boldsymbol{H}_a) \subseteq \boldsymbol{K}(a)$, along with a morphism $u : \mathbf{1}_a \to f$ in $\boldsymbol{K}(a)$. We show that $\boldsymbol{K}$ is isomorphic to $\boldsymbol{K}_f$, by a unique isomorphism.

To do that, let $c \in \boldsymbol{G}_0$ be an arbitrary object, and let v be a morphism of $\boldsymbol{K}_f(c) = \underline{\mathbf{Hom}}^L(c, f)$ (so a 2-morphism of $\boldsymbol{G}$). Denote by $g \in \underline{\mathbf{Hom}}^L(c, a)_0$ the domain of v. Then v can also be viewed as a morphism in $\underline{\mathbf{Hom}}^L(g, b)$, and on the other hand, we have the canonical morphism w from g to $\mathbf{1}_a$ in $\underline{\mathbf{Hom}}^L(g, a)$. Applying $t_{g,\boldsymbol{K}}$ to the morphism $u \otimes w \in \boldsymbol{K}(a) \circ_{\underline{\mathbf{Hom}}^L(a,a)} \underline{\mathbf{Hom}}^L(g, a)$, we may write $t_{g,\boldsymbol{K}}(u \otimes w)$ as $v \otimes x$ for a unique $x \in \boldsymbol{K}(c)$, which we take to be the image of v. By construction this map commutes with the structure maps t, and is unique with this property.

$\underline{k=2}$: We need to show that each isomorphism $\alpha : \boldsymbol{K}_g \circ \boldsymbol{K}_f \to \boldsymbol{K}_h$ with $f : a \to b$, $g : b \to c$ and $h : a \to c$ arises from a unique $\gamma : D^2 \to \boldsymbol{G}$, with boundary f, g, h (as in the end of Example 3.3.8). Uniqueness was already shown in the part $k = 1$. For existence, we apply $\alpha_b : \boldsymbol{K}_g(b) \circ \boldsymbol{K}_f(b) \to \boldsymbol{K}_h(b)$ to the element $\mathbf{1}_g \otimes \mathbf{1}_f$ (where $\mathbf{1}_g$ is the identity morphism of the object g of $\underline{\mathbf{Hom}}^L(b, c)$, viewed as an element of $\boldsymbol{G}_2$, and similarly for f), to obtain an element $\gamma_b \in \boldsymbol{K}_h(b)_1$, again viewed as a 2-morphism of $\boldsymbol{G}$. It is clear that the map α coincides with α_γ on the given maps, and then, again by the uniqueness statement, that $\alpha = \alpha_\gamma$ globally. $\square$

3.3.10. *Recovering a definable* 2*-groupoid.* The main statement of classical internality starts with the assumption of internality, and produces a definable (nonempty) groupoid from it. The general outline of this construction was recalled in Section 2.2, and in Proposition 2.2.6 we indicated how this construction is useful in the description of definable sets in the cover.

In our approach, the construction of the (2-)groupoid is almost tautological: we defined a groupoid (or a 2-groupoid) associated to every stable expansion, and by definition, the expansion is an internal cover if the groupoid is nonempty. However, we still need to show that the groupoid is equivalent to a definable one, which we sketch below. The other part, describing the (admissible) 1-groupoids in the cover in terms of suitable definable fibrations in the base, is more involved, and we postpone most of the work here to future work.

Proposition 3.3.11. *Let $\mathcal{T}^*$ be a 2-internal cover of $\mathcal{T}$. Then the 2-groupoid $I^2_{\mathcal{T}^*/\mathcal{T}}$ associated to it is equivalent to a $\mathcal{T}$-definable one.*

Proof sketch. Assume $\mathcal{T}^*$ is a 2-internal cover of $\mathcal{T}$, and let $\omega : \mathcal{T}^* \to \mathcal{T}_A$ be a 2-stable interpretation. For simplicity we assume that Γ, the collection of admissible covers, consists of one definable family. As in 3.3.6, we have a fixed parameter u_0 and a uniform family $\boldsymbol{K} = \boldsymbol{K}_c$ of groupoids in $\mathcal{T}^*$ defined over u_0, along with (uniformly definable) weak equivalences $f_c : \boldsymbol{X}_c \overset{\sim}{\dashrightarrow} \boldsymbol{K}_c$ and $g_c : \omega(\boldsymbol{X}_c) \to \boldsymbol{K}_c$ for $\boldsymbol{X}_c$ members of Γ (note that c ranges over a definable set in $\mathcal{T}$ by assumption). Like in Remark 3.3.7, as u_0 varies, we obtain a family $\boldsymbol{K}_{u,c}$ of objects of a definable 1-groupoid $\boldsymbol{P}_c$, along with a map $\boldsymbol{P}_c \to \underline{\mathbf{Iso}}(\boldsymbol{X}, \omega(\boldsymbol{X}))$ for each member $\boldsymbol{X}$ of Γ. Furthermore, $\boldsymbol{P}_c$ itself is also in Γ. Applying the above map to $\boldsymbol{X} = \boldsymbol{P}_{c'}$, we obtain a family of definable maps of 1-groupoids $a : \boldsymbol{P}_c \to \underline{\mathbf{Iso}}(\boldsymbol{P}_{c'}, \omega_c(\boldsymbol{P}_{c'}))$.

The 2-groupoid $\boldsymbol{G}$ is constructed as follows: $\boldsymbol{G}_0$ is the definable set of parameters c as above. Each groupoid $\boldsymbol{P}_c$ will be isomorphic to $\underline{\mathbf{Hom}}^L(*, c)$ in the corresponding $\boldsymbol{G}^*$. Let c, d be two elements of $\boldsymbol{G}_0$. Given an object u of $\boldsymbol{P}_c$, the map a above produces a groupoid $\boldsymbol{K}_u$ as an object of $\underline{\mathbf{Iso}}(\boldsymbol{P}_d, \omega_c(\boldsymbol{P}_d))$, along with weak equivalences $f : \boldsymbol{P}_d \overset{\sim}{\dashrightarrow} \boldsymbol{K}_u$ and $g : \omega_c(\boldsymbol{P}_d) \to \boldsymbol{K}_u$. Let $\boldsymbol{Q}_{u,v}$ be the set of morphisms in $\boldsymbol{K}_u$ with domain $f(v)$. We set the morphisms from c to d to be the definable types space of $\boldsymbol{Q}_{u,v}$ over $\mathcal{T}$ (this is definable in $\mathcal{T}$ by stability of the embedding). Note that each such type includes, in particular, the information of the object of $\omega_c(\boldsymbol{P}_d)$ which is the codomain of any realisation (as an element of $\boldsymbol{Q}_{u,v}$).

Similarly, assume e is another element of $\boldsymbol{G}_0$, and w an object of $\boldsymbol{P}_e$. The elements of $\boldsymbol{G}_2$ with vertices c, d, e are defined as the types over $\mathcal{T}$ of triples $\boldsymbol{Q}_{u,v} \times \boldsymbol{Q}_{v,w} \times \boldsymbol{Q}_{u,w}$ (over all such u, v, w). The 2-composition is defined similarly, by considering 4-tuples. We skip the details of the construction, as well as the proof that the map determined by a is a weak equivalence. $\square$

We mention also that this description can, in principle, be used to give an equivalent combinatorial definition of 2-internality (similar to the original definition of internality), but I could not find one sufficiently pleasant to write.

3.4. ***Questions.*** I mention a few natural questions that I hope to address in the future.

3.4.1. *Structure of admissible internal covers.* The definable version of the 2-groupoid $\boldsymbol{G}$ associated to a 2-internal cover $\mathcal{T}^*$ of $\mathcal{T}$ was only sketched above. Assuming it is properly described, it still needs to be seen that $\mathcal{T}_{\boldsymbol{G}}$ and $\mathcal{T}^*$ are, in some sense, equivalent. It would also be useful to describe the admissible covers (in either) as suitably defined "higher local systems" on $\boldsymbol{G}$. Both questions require that we name the precise closure properties on the collection of admissible covers: we already assumed that they are closed under finite inverse limit and definable mapping spaces, but it is not clear, for example, if some closure under quantifiers is required.

3.4.2. *Lax interpretations.* We had not run into the questions above because we required objects of the 2-groupoid I^2 to be actual interpretations. This works well in the example of $\mathcal{T}_{\boldsymbol{G}}$, but for general expansions it might make more sense to consider a larger class of "lax interpretations" that preserve only the admissible covers (possibly up to weak equivalence).

3.4.3. *Internal covers of $\mathcal{T}_{\boldsymbol{G}}$.* The requirement for introducing admissible covers was motivated above. However, in the case of $\mathcal{T}_{\boldsymbol{G}}$ it might still be true that essentially all internal covers are the ones described (up to covers that come from the base $\mathcal{T}$). Again, stating this precisely requires clarifying the structure of the collection of admissible covers.

3.4.4. *Relation to analysability.* The 2-groupoid $\boldsymbol{G}^*$ in the theory $\mathcal{T}$ provides an example of a 2-analysable set over $\mathcal{T}$. Can we describe (combinatorially) which 2-analysable covers occur in this way?

3.4.5. *Generalisation to higher dimensions.* This is rather clear: one continues by induction, defining an $(i+1)$-groupoid associated to a stable expansion by taking into account i-internal covers, and then defining the expansion to be an $(i+1)$-cover if this groupoid is nonempty. However, some of the proofs given above would be difficult to generalise, and it would be interesting to look for a smoother way. In any case, this only applies to each finite level, and it does not seem reasonable to expect a generalisation to arbitrary ∞-groupoids.

3.4.6. *Structure at $*$.* We did not consider the structure of the groupoid $\underline{\mathbf{Hom}}^L_{\boldsymbol{G}^*}(*,*)$ definable in $\mathcal{T}_{\boldsymbol{G}}$. On top of the groupoid structure, composition gives it a structure of a monoidal category up to homotopy (i.e., the homotopy category is monoidal). It also acts on all the admissible covers, so it is really a higher analogue of the binding group. However, we did not consider what could be a version of the Galois correspondence or of descent, as in the 1-dimensional case.

Acknowledgement and dedication

I am delighted to thank Tomer Schlank's gang at the Hebrew University (aka "Ha'arakya"), and particularly Asaf Horev, Shaul Barkan, Shai Keidar and Tomer himself for answering what must have been rather bizarre and elementary questions in homotopy theory. They form a very helpful and fun community.

This paper is a late expansion on some vague ideas that I presented in the postponed online conference that honoured Ehud Hrushovski for his 60th birthday. It is a pleasure to thank Udi again for his guidance and dedicate the paper (hopefully clearer than my presentation at that talk!) to him.

References

[1] J. Goodrick, B. Kim, and A. Kolesnikov, "Type-amalgamation properties and polygroupoids in stable theories", *J. Math. Log.* **15**:1 (2015), art. id. 1550004. MR Zbl

[2] L. Haykazyan and R. Moosa, "Functoriality and uniformity in Hrushovski's groupoid-cover correspondence", *Ann. Pure Appl. Logic* **169**:8 (2018), 705–730. MR Zbl

[3] E. Hrushovski, "Computing the Galois group of a linear differential equation", pp. 97–138 in *Differential Galois theory* (Będlewo, Poland, 2001), edited by T. Crespo and Z. Hajto, Banach Center Publ. **58**, Polish Acad. Sci. Inst. Math., Warsaw, 2002. MR Zbl

[4] E. Hrushovski, "Groupoids, imaginaries and internal covers", *Turkish J. Math.* **36**:2 (2012), 173–198. MR Zbl

[5] A. Joyal, "Quasi-categories and Kan complexes", *J. Pure Appl. Algebra* **175**:1-3 (2002), 207–222. MR Zbl

[6] M. Kamensky, "A categorical approach to internality", pp. 139–156 in *Models, logics, and higher-dimensional categories* (Montréal, 2009), edited by B. Hart et al., CRM Proc. Lecture Notes **53**, Amer. Math. Soc., Providence, RI, 2011. MR Zbl

[7] M. Kamensky and A. Pillay, "Interpretations and differential Galois extensions", *Int. Math. Res. Not.* **2016**:24 (2016), 7390–7413. MR Zbl

[8] J. Lurie, *Higher topos theory*, Annals of Math. Studies **170**, Princeton University Press, 2009. MR Zbl

[9] J. Lurie, "Kerodon", electronic reference, 2018, available at https://kerodon.net.

[10] M. Makkai and G. E. Reyes, *First order categorical logic*, Springer, 1977. MR Zbl

[11] B. Poizat, "Une théorie de Galois imaginaire", *J. Symbolic Logic* **48**:4 (1983), 1151–1170. MR Zbl

[12] P. Wang, "Extending Hrushovski's groupoid-cover correspondence using simplicial groupoids", *Ann. Pure Appl. Logic* **172**:7 (2021), art. id. 102970. MR Zbl

[13] 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., Lect. Notes in Math. **834**, Springer, 1980. MR Zbl

Received 9 Oct 2022. Revised 6 Mar 2023.

MOSHE KAMENSKY:
kamensky.bgu@gmail.com
Department of Mathematics, Ben-Gurion University, Be'er-Sheva, Israel

msp

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. 2 2023

Celebratory issue on the occasion of
Ehud Hrushovski's 60th Birthday