Pacific Journal of Mathematics

Volume 271 No. 1 September 2014

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

Founded in 1951 by E. F. Beckenbach (1906–1982) and F. Wolf (1904–1989)

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

The subscription price for 2014 is US $410/year for the electronic version, and $535/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and Web of Knowledge (Science Citation Index).

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published twelve times a year. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163.

PJM peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.1

PROPER HOLOMORPHIC MAPS BETWEEN BOUNDED SYMMETRIC DOMAINS REVISITED

GAUTAM BHARALI AND JAIKRISHNAN JANARDHANAN

We prove that a proper holomorphic map between two nonplanar bounded symmetric domains of the same dimension, one of them being irreducible, is a biholomorphism. Our methods allow us to give a single, all-encompassing argument that unifies the various special cases in which this result is known. We discuss an application of these methods to domains having noncompact automorphism groups that are not assumed to act transitively.

1. Introduction and statement of results

The primary objective of this paper is to prove the following result:

Theorem 1.1. *Let D_1 and D_2 be two bounded symmetric domains of complex dimension $n \geq 2$. Assume that either D_1 or D_2 is irreducible. Then, any proper holomorphic mapping of D_1 into D_2 is a biholomorphism.*

This theorem is known in several special cases. For $D_1 = D_2 = \mathbb{B}^n$, the (Euclidean) ball in $\mathbb{C}^n$ with $n \geq 2$, the result was established by Alexander [1977]. This is a pioneering work that has motivated several generalizations to proper holomorphic maps between certain types of smoothly bounded pseudoconvex domains. Henkin and Novikov [1984] described a method for proving the result when $D_1 = D_2 =: D$ where D is a bounded symmetric domain that is *not* of tube type. Later, Tsai [1993] established the result for D_1 and D_2 as above, provided D_1 is irreducible and rank $D_1 \geq$ rank $D_2 \geq 2$.

Tsai's result is a broad metric-rigidity theorem (under the Bergman metric) for proper holomorphic maps of D_1 into D_2, where D_1 and D_2 are as above but not necessarily equidimensional. In such a result, the condition rank $D_1 \geq$ rank $D_2 \geq 2$ is indispensable. Adapting Tsai's ideas to the equidimensional case, Tu [2002] established Theorem 1.1 in the higher-rank case, assuming D_1 *is irreducible*. In using Tsai's ideas, the assumption that D_1 is irreducible is essential — see [Tu 2002,

Bharali is supported by a UGC Centre for Advanced Study grant. Janardhanan is supported by a UGC Centre for Advanced Study grant and by a scholarship from the IISc.
MSC2010: primary 32H02, 32M15; secondary 32H40.
Keywords: bounded symmetric domains, Harish-Chandra realization, Jordan triple systems, proper holomorphic maps, rigidity, Schwarz lemma.

Proposition 3.3] — and it is not clear that a small mutation of those ideas allows one to weaken this assumption. In our work, we are able to assume either D_1 or D_2 to be irreducible precisely by not relying too heavily on the fine structure of these domains. Indeed, we wish to emphasize that the focus of this work is *not* on mopping up the residual cases in Theorem 1.1. The methods in [Henkin and Novikov 1984] (which rely on [Tumanov and Khenkin 1982]), [Tsai 1993] and [Tu 2002] are tied, in a rather maximalistic way, to the fine structure of a bounded symmetric domain. In contrast, we present some ideas that make very mild use of the underlying orbit structure of the bounded symmetric domains. They could therefore be applied to manifolds whose automorphism groups are not assumed to act transitively but are merely "large enough". Theorem 1.5 is an illustration of this notion. These ideas also provide a *unified* argument, irrespective of rank or reducibility, for Theorem 1.1.

We need to be more precise about the preceding remarks. This requires some elaboration on the objects of interest. A bounded symmetric domain in $\mathbb{C}^n$ is the holomorphic imbedding in $\mathbb{C}^n$ of some Hermitian symmetric space of noncompact type. It is *irreducible* if it is not a product of bounded symmetric domains of lower dimension. Cartan studied Hermitian symmetric spaces of noncompact type and classified the irreducible ones, showing that they are one of six types of homogeneous spaces. An outcome of [Harish-Chandra 1956] is that these homogeneous spaces (and products thereof) can be imbedded in $\mathbb{C}^n$ as bounded convex balanced domains (we say that a domain $D \subset \mathbb{C}^n$ is *balanced* if, for any $z \in D$, we have $\zeta z \in D$ for each ζ in the closed unit disc centered at $0 \in \mathbb{C}$). This imbedding is unique up to a linear isomorphism of $\mathbb{C}^n$. Such a realization of a bounded symmetric domain is called a *Harish-Chandra realization*.

The three main features that we wish to emphasize about this work are:

(a) The arguments in [Alexander 1977] involve estimates showing how a proper mapping maps conical regions with vertex on $\partial\mathbb{B}^n$ into the "admissible" approach regions of Korányi [1969]. Boundary approach, in a somewhat different sense, plus Chern–Moser theory [1974] make an appearance in [Tumanov and Khenkin 1982]. In contrast, apart from, and owing to, a result of Bell [1982] on boundary behavior, our proof involves rather "soft" methods.

(b) The techniques underlying [Tsai 1993] and [Tu 2002] rely almost entirely on the fine structure of a bounded symmetric domain. Specifically, they involve studying the effect of a proper holomorphic map on the characteristic symmetric subspaces of a bounded symmetric domain of rank at least 2. In contrast, our techniques rely on only a coarse distinction between the different strata that comprise the boundary of an irreducible bounded symmetric domain (e.g., see Remark 4.8 below).

(c) An advantage of arguments that rely on only a coarse resolution of a bounded symmetric domain is that some of them are potentially applicable to the study of domains that have noncompact automorphism groups but are not assumed to be symmetric. A demonstration this viewpoint is the proof of Theorem 1.5 below.

Let D be a bounded symmetric domain. The main technical tool that facilitates our study of the structure of ∂D, and describes certain elements of $\mathrm{Aut}(D)$ with the optimal degree of explicitness, is the notion of Jordan triple systems. The application of Jordan triple systems to geometry appears to have been pioneered by Koecher [1999]. Our reference on this subject is the lecture notes of Loos [1977], which are devoted specifically to bounded symmetric domains. Jordan triple systems and versions of the Schwarz lemma are our primary tools. We present next an outline of how we use these tools.

An important lemma, which is inspired by Alexander's work, is the following.

Key Lemma 1.2. *Let D be a realization of an irreducible bounded symmetric domain of dimension $n \geq 2$ as a bounded convex balanced domain in $\mathbb{C}^n$. For $z \in D \setminus \{0\}$, let $\Delta_z := \{\zeta z : \zeta \in \mathbb{C}$ and $\zeta z \in D\}$. Let W_1 and W_2 be two regions in D such that $0 \in W_1 \cap W_2$ and let $F : D \to D$ be a holomorphic map. Assume that*:

(i) *F maps W_1 biholomorphically onto W_2 with $F(0) = 0$.*

(ii) *There exists a nonempty open set $U \subset W_1 \setminus \{0\}$ such that, for each $z \in U$, $\Delta_z \subset W_1$ and $\Delta_{F(z)} \subset W_2$.*

Then, F is an automorphism of D.

This is a consequence of Vigué's Schwarz lemma [1991] (see Result 4.6 below), and the irreducibility of D is essential to the lemma.

Our proof of Theorem 1.1 may be summarized as follows (we will assume here that D_1 and D_2 are Harish-Chandra realizations of the domains in question):

- By Bell's theorem [1982, Theorem 2], F extends to a neighborhood of $\overline{D}_1$ and we can find a point p in the Bergman–Shilov boundary of D_1, and a small ball B around it, such that $F|_B$ is a biholomorphism.
- We may assume that $F(0) = 0$. Let $\{a_k\}$ be a sequence in $D_1 \cap B$ converging to p and let $b_k := F(a_k)$. Let $\phi_k^j \in \mathrm{Aut}(D_j)$ be an automorphism that maps 0 to a_k if $j = 1$ and to b_k if $j = 2$. It turns out that both p and $F(p)$ are peak points, whence $\phi_k^j \to p^{(j)}$ uniformly on compact subsets, where $p^{(1)} := p$ and $p^{(2)} := F(p)$.
- Using the Schwarz lemma for convex balanced domains (Result 4.5 below) we show that a subsequence of $\{(\phi_j^2)^{-1} \circ F \circ \phi_j^1\}$ converges to a linear map and that, owing to the tautness of D_1 and D_2, this map is a biholomorphism of D_1 onto D_2.

- We may now take $D_1 = D_2 = D$. We shall use our Key Lemma 1.2 with $W_1 = (\phi_k^1)^{-1}(D \cap B)$ and $W_2 = (\phi_k^2)^{-1}(D \cap F(B))$ for k sufficiently large.
- Since the analytic discs Δ_z and $\Delta_{F(z)}$ are not relatively compact in D, the mode of convergence of $\{\phi_k^j\}$ isn't a priori good enough to infer that appropriate families of these discs will be swallowed up by W_j, $j = 1, 2$. By Bell's theorem, each ϕ_k^j extends to some neighborhood of D. We show that $\{\phi_k^j\}$, passing to a subsequence and relabeling if necessary, converges *uniformly* on certain special circular subsets of D that are adherent to ∂D. This is enough to overcome the difficulty just described.

Let us define a term that we used in the sketch above, which we shall also need in stating our next theorem.

Definition 1.3. Let $D \subsetneq \mathbb{C}^n$ be a domain and let $p \in \partial D$. We say that p is a *peak point* if there exists a function $h \in \mathcal{O}(D) \cap \mathcal{C}(\overline{D}; \mathbb{C})$ such that $h(p) = 1$ and $|h(z)| < 1$ for all $z \in \overline{D} \setminus \{p\}$. The function h is called a *peak function for* p.

When a domain D is bounded, the noncompactness of $\mathrm{Aut}(D)$ (in the compact-open topology) is equivalent to D having a boundary orbit-accumulation point; see [Narasimhan 1971].

Definition 1.4. Let $D \subsetneq \mathbb{C}^n$ be a domain and let $p \in \partial D$. We say that p is a *boundary orbit-accumulation point* if there exist a point $a \in D$ and a sequence of automorphisms $\{\phi_k\}$ of D such that $\lim_{k\to\infty} \phi_k(a) = p$.

With the last two definitions, we are in a position to state our second theorem. Note that D_1 is *not* assumed to be a bounded symmetric domain. Yet, some of the techniques sketched above (versions of which have been used to remarkable effect in the literature in this field) are general enough to be applicable to the following situation.

Theorem 1.5. *Let D_1 be a bounded convex balanced domain in $\mathbb{C}^n$ whose automorphism group is noncompact and let p be a boundary orbit-accumulation point. Let D_2 be a realization of a bounded symmetric domain as a bounded convex balanced domain in $\mathbb{C}^n$. Assume that there is a neighborhood U of p and a biholomorphic map $F : U \to \mathbb{C}^n$ such that $F(U \cap D_1) \subset D_2$ and $F(U \cap \partial D_1) \subset \partial D_2$. Assume that either p or $F(p)$ is a peak point. Then, there exists a* linear *map that maps D_1 biholomorphically onto D_2.*

Remark 1.6. Theorem 1.5 (together with Bell's theorem [1982]) gives a very short proof of the rigidity theorem of Mok and Tsai [1992] under the additional assumption that the convex domain D in their result is also circular. There is extensive literature on rigidity theorems relating to bounded symmetric domains, but we shall not dwell any further on it.

Remark 1.7. A version of this result can be proved without assuming that D_1 is either balanced or convex. D_1 merely needs to be complete Kobayashi hyperbolic. However, in this case, the biholomorphism of D_1 onto D_2 will not, in general, be linear. We prefer the above version: the conclusion that there exists a *linear* equivalence places Theorem 1.5 among the rigidity theorems alluded to in Remark 1.6.

The layout of this paper is as follows. Since Jordan triple systems play a vital role in describing not just the structure of the boundary of a bounded symmetric domain, but also some of its key automorphisms, we begin with a primer on Jordan triple systems. Readers who are familiar with Jordan triple systems can skip to Section 3, where we discuss the boundary geometry of bounded symmetric domains. Section 4 is devoted to stating and proving certain propositions that are essential to our proofs. Finally, in Sections 5 and 6, we present the proofs of the results stated above.

2. A primer on Jordan triple systems

There is a natural connection between bounded symmetric domains and certain Hermitian Jordan triple systems. This section collects several definitions and results that are required to give a coherent description of the boundary of a bounded symmetric domain (which we shall discuss in the next section).

Unless otherwise stated, the results in this section can be found in the UC Irvine lectures by Loos [1977] describing how Jordan triple systems can be used to study the geometry of bounded symmetric domains.

Definition 2.1. A *Hermitian Jordan triple system* is a complex vector space V endowed with a triple product $(x, y, z) \mapsto \{x, y, z\}$ that is symmetric and bilinear in x and z and conjugate-linear in y and satisfies the Jordan identity

$$\{x, y, \{u, v, w\}\} - \{u, v, \{x, y, w\}\} = \{\{x, y, u\}, v, w\} - \{u, \{y, x, v\}, w\} \qquad \text{for all } x, y, u, v, w \in V.$$

Such a system is said to be *positive* if for each $x \in V \setminus \{0\}$ for which $\{x, x, x\} = \lambda x$ (where λ is a scalar), we have $\lambda > 0$.

As mentioned in Section 1, a bounded symmetric domain of complex dimension n has a realization D as a bounded convex balanced domain in $\mathbb{C}^n$. Let $(z_1, \dots, z_n)$ be the global holomorphic coordinates coming from the product structure on $\mathbb{C}^n$ and let $(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n)$ denote the standard ordered basis of $\mathbb{C}^n$. Let K_D denote the Bergman kernel of (the above realization of) D and h_D the Bergman metric. The function $\{\cdot, \cdot, \cdot\} : \mathbb{C}^n \times \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}^n$ obtained by the requirement

$$h_D(\{\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_k\}, \boldsymbol{\epsilon}_l) = \left. \frac{\partial^4 \log K_D(z, z)}{\partial z_i \, \partial \bar{z}_j \, \partial z_k \, \partial \bar{z}_l} \right|_{z=0}, \tag{2-1}$$

and by extending $\mathbb{C}$-linearly in the first and third variables and $\mathbb{C}$-antilinearly in the second, has the property that $(\mathbb{C}^n, \{\cdot\,,\cdot\,,\cdot\})$ is a positive Hermitian Jordan triple system (abbreviated hereafter as PHJTS). This relationship is a one-to-one correspondence between finite-dimensional PHJTSs and bounded symmetric domains — which we shall make more precise in Section 3.

Let $(V, \{\cdot\,,\cdot\,,\cdot\})$ be an HJTS. It will be convenient to work with the operators

$$\mathsf{D}(x, y)z = \mathsf{Q}(x, z)y := \{x, y, z\}. \tag{2-2}$$

We define the operator $Q : V \to \mathrm{End}(V)$ by $Q(x)y := \mathsf{Q}(x, x)y/2$. For any $x \in V$, we can define the so-called *odd powers* of x recursively by

$$x^{(1)} := x \qquad \text{and} \qquad x^{(2p+1)} := Q(x)x^{(2p-1)} \quad \text{if } p \geq 1.$$

A vector $e \in V$ is called a tripotent if $e^{(3)} = e$.

Tripotents are important to this discussion because:

- A finite-dimensional PHJTS has plenty of nonzero tripotents.
- Given a finite-dimensional PHJTS $(V, \{\cdot\,,\cdot\,,\cdot\})$, any vector V has a certain canonical decomposition as a linear combination of tripotents.
- In a finite-dimensional PHJTS, the set of tripotents forms a real-analytic submanifold.

We refer the interested reader to [Loos 1977, Chapter 3] for details of the first fact. As for the second fact, we need a couple of new notions. First, given an HJTS $(V, \{\cdot\,,\cdot\,,\cdot\})$, we say that two tripotents $e_1, e_2 \in V$ are orthogonal if $\mathsf{D}(e_1, e_2) = 0$. Second, given $x \in V$, we define the real vector space $\ll x \gg$ by

$$\ll x \gg := \mathrm{span}_{\mathbb{R}}\{x^{(2p+1)} : p = 0, 1, 2, \dots\}.$$

These two notions allow us to state the following:

Result 2.2 (spectral decomposition theorem). *Let $(V, \{\cdot\,,\cdot\,,\cdot\})$ be a finite-dimensional PHJTS. Then, each $x \in V \setminus \{0\}$ can be written uniquely as*

$$x = \lambda_1 e_1 + \cdots + \lambda_s e_s, \tag{2-3}$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_s > 0$ and $\{e_1, \dots, e_s\}$ is an $\mathbb{R}$-basis of $\ll x \gg$ comprising pairwise orthogonal tripotents.

The decomposition of $x \in V$ as given by Result 2.2 is called the *spectral decomposition* of x. The assignment $x \mapsto \lambda_1(x)$, where $\lambda_1(x)$ is as given by (2-3), is a well-defined function and can be shown to be a norm on V. This norm is called the *spectral norm* on V.

Next, we present another decomposition, which gives us the second ingredient needed to describe the boundary geometry of a bounded symmetric domain.

Result 2.3 (Pierce decomposition). *Let* $(V, \{\cdot\,, \cdot\,, \cdot\})$ *be an HJTS and let* $e \in V$ *be a tripotent. Then, the spectrum of* $\mathsf{D}(e, e)$ *is a subset of* $\{0, 1, 2\}$. *Let*

$$V_j = V_j(e) := \{x \in V : \mathsf{D}(e, e)x = jx\}, \quad j \in \mathbb{Z}.$$

Then:

(a) $V = V_0 \oplus V_1 \oplus V_2$.

(b) *If* $e \neq 0$, *then* $e \in V_2$.

(c) *We have the relation* $\{V_\alpha, V_\beta, V_\gamma\} \subset V_{\alpha-\beta+\gamma}$.

(d) V_0, V_1 *and* V_2 *are Hermitian Jordan subsystems of* $\{\cdot\,, \cdot\,, \cdot\}$.

The direct-sum decomposition (a) given by this result is called the *Pierce decomposition of* V *with respect to the tripotent* e. The ideas that go into proving the Pierce decomposition theorem allow us to construct a special partial order on the set of tripotents of V. In order to avoid statements that are vacuously true, unless stated otherwise, we take $(V, \{\cdot\,, \cdot\,, \cdot\})$ to be a PHJTS. Let $e, e' \in V$ be tripotents. We say that *e is dominated by e'* ($e \preceq e'$) if there is a tripotent e_1 orthogonal to e such that $e' = e + e_1$. We say that *e is strongly dominated by e'* ($e \prec e'$) if $e \preceq e'$ and $e \neq e'$. The result of interest, in this regard, is the following:

Result 2.4. *Let* $(V, \{\cdot\,, \cdot\,, \cdot\})$ *be an HJTS. Let* $e_1, e_2 \in V$ *be orthogonal tripotents and let* $e = e_1 + e_2$. *If* $e' \in V$ *is a tripotent orthogonal to* e, *then* e' *is orthogonal to* e_1 *and* e_2.

Now suppose $\{\cdot\,, \cdot\,, \cdot\}$ *is positive. Then, the relation* $\preceq$ *is a partial order on the set of tripotents.*

Definition 2.5. A tripotent is said to be *minimal* (or *primitive*) if it is minimal for $\preceq$ among nonzero tripotents. It is said to be *maximal* if it is maximal for $\preceq$.

Result 2.6. *Consider the tripotents of* V *partially ordered by* $\preceq$.

(1) *A tripotent* e *is maximal if and only if the Pierce space* $V_0(e)$ *equals* 0.

(2) *If, for a tripotent* e, *the Pierce space* $V_2(e)$ *equals* $\mathbb{C}e$, *then* e *is primitive.*

Let us now also assume that $(V, \{\cdot\,, \cdot\,, \cdot\})$ is finite-dimensional. Given any nonzero tripotent e, it follows from finite-dimensionality and the repeated application of Result 2.4 that e can be written as a sum of mutually orthogonal primitive tripotents. This brings us to the final concept in this primer: the *rank of a tripotent e* is the minimum number of primitive tripotents required for such a decomposition of e, while the *rank of* $(V, \{\cdot\,, \cdot\,, \cdot\})$ is the highest rank that a tripotent of V can have.

3. The boundary geometry of bounded symmetric domains

In this section we describe the boundary of a bounded symmetric domain in terms of the positive Hermitian Jordan triple system associated to it. Thus, we shall follow the notation introduced in Section 2. Recall that a bounded symmetric domain D has a realization as a bounded convex balanced domain. When we say "Hermitian Jordan triple system associated to D", it is implicit that D is this realization and the association is the one given by (2-1). This is a one-to-one correspondence, described as follows:

Result 3.1 [Loos 1977, Theorem 4.1]. *Let D be a realization of a bounded symmetric domain as a bounded convex balanced domain in $\mathbb{C}^n$ for some $n \in \mathbb{Z}_+$. Then, D is the open unit ball in $\mathbb{C}^n$ with respect to the spectral norm determined by the PHJTS associated to D. Conversely, given a PHJTS* $(\mathbb{C}^n, \{\cdot, \cdot, \cdot\})$, *the open unit ball with respect to the spectral norm determined by it is a bounded symmetric domain D, and the PHJTS associated to D by the rule* (2-1) *is* $(\mathbb{C}^n, \{\cdot, \cdot, \cdot\})$.

In what follows, whenever we mention a bounded symmetric domain D, *it will be understood that D is a bounded convex balanced realization.*

The boundary of a bounded symmetric domain $D \subset \mathbb{C}^n$ has a certain stratification into real-analytic submanifolds that can be described in terms of the PHJTS associated to D. The first part of this section is devoted to describing this stratification. Fix a bounded symmetric domain $D \subset \mathbb{C}^n$ and let $(\mathbb{C}^n, \{\cdot, \cdot, \cdot\}_D)$ be the PHJTS associated to it. It turns out (see [Loos 1977, Theorem 5.6]) that the set M_D of tripotents of $\mathbb{C}^n$ with respect to $\{\cdot, \cdot, \cdot\}_D$ is a disjoint union of real-analytic submanifolds of $\mathbb{C}^n$. For each $e \in M_D$, let $M_{D,e}$ denote the connected component of M_D containing e. The tangent space $T_e(M_{D,e})$, viewed *extrinsically* (i.e., so that $e + T_e(M_{D,e})$ is the affine subspace of all tangents to $M_{D,e}$ at e), is

$$T_e(M_{D,e}) = i\,A(e) \oplus V_1(e),$$

where $A(e)$ is determined by the relation $V_2(e) = \{x + iy \in \mathbb{C}^n : x, y \in A(e)\}$ and $V_j(e)$ is the eigenspace of $j = 0, 1, 2$ in the Pierce decomposition of $\mathbb{C}^n$ with respect to e.

Let M_D^* be the set of all nonzero tripotents and let $\|\cdot\|_D$ denote the spectral norm determined by $\{\cdot, \cdot, \cdot\}_D$. Define

$$E_D := \{(e, v) \in \mathbb{C}^n \times \mathbb{C}^n : e \in M_D^* \text{ and } v \in V_0(e)\},$$
$$\mathfrak{B}_D := \{(e, v) \in E_D : \|v\|_D < 1\}.$$

We can write $\mathfrak{B}_D$ as a disjoint union of the form

$$\mathfrak{B}_D := \bigsqcup_{\alpha \in \mathscr{C}} \mathfrak{B}_{D,\alpha}, \tag{3-1}$$

where $\mathscr{C}$ is the set of connected components of M^*_D, and each $\mathfrak{B}_{D,\alpha}$ is a connected, real-analytic submanifold of $\mathbb{C}^n \times \mathbb{C}^n$ that is a real-analytic fiber bundle whose fibers are unit $\|\cdot\|_D$-discs. The key theorem about the boundary of D is as follows:

Result 3.2 [Loos 1977, Chapter 6]. *Let D be a bounded symmetric domain in $\mathbb{C}^n$ and let $\mathsf{f} : \mathfrak{B}_D \to \mathbb{C}^n$ be defined by $\mathsf{f}(e, v) := e + v$. Then*:

(i) $\mathsf{f}|_{\mathfrak{B}_{D,\alpha}}$ *is an imbedding for each $\alpha \in \mathscr{C}$.*

(ii) $\partial D = \bigsqcup_{\alpha \in \mathscr{C}} \mathscr{M}_{D,\alpha}$, *where $\mathscr{M}_{D,\alpha} := \mathsf{f}(\mathfrak{B}_{D,\alpha})$.*

(iii) *In the above stratification of ∂D, if $\mathscr{M}_{D,\alpha}$ is of dimension d_α, then it is a closed, connected, real-analytic imbedded submanifold of the open set*

$$\mathbb{C}^n \setminus \bigcup_{\beta\,:\,\dim_{\mathbb{R}}(\mathscr{M}_{D,\beta}) < d_\alpha} \mathscr{M}_{D,\beta}.$$

Furthermore, when D is an *irreducible* bounded symmetric domain in $\mathbb{C}^n$, then we can provide further information. Here, the rank of a bounded symmetric domain is the rank of the Jordan triple system $(\mathbb{C}^n, \{\cdot, \cdot, \cdot\}_D)$.

Result 3.3 [Loos 1977, Chapter 6; Vigué 1991, Théorème 7.3]. *Let D be an irreducible bounded symmetric domain in $\mathbb{C}^n$ of rank r, and let $\mathscr{C}$ denote the set of connected components of $\mathfrak{B}_D$. Then*:

(i) *$\mathscr{C}$ has cardinality r.*

(ii) *Each connected component of the decomposition* (3-1) *is a bundle over a submanifold of nonzero tripotents of rank j, $j \in \{1, \dots, r\}$. Denoting this bundle as $\mathfrak{B}_{D,j}$, $j \in \{1, 2, \dots, r\}$, we can express the stratification of ∂D given by Result 3.2 (ii) as*

$$\partial D = \bigsqcup_{j=1}^{r} \mathscr{M}_{D,j},$$

where $\mathscr{M}_{D,j} := \mathsf{f}(\mathfrak{B}_{D,j})$, and each $\mathscr{M}_{D,j}$ is connected.

(iii) *The stratum $\mathscr{M}_{D,1}$ is dense in ∂D.*

The other goal of this section is to describe the structure of the germs of complex-analytic varieties contained in the boundary of a bounded symmetric domain D. This structure can be described in extremely minute detail; see, for instance, [Wolf 1972]. In fact, the papers about higher-rank bounded symmetric domains mentioned in Section 1 make extensive use of this fine structure. However, *in this work, we only need very coarse information about the complex analytic structure of ∂D*, specifically, the distinction between the Bergman–Shilov boundary of D and its complement in ∂D.

We denote the Bergman–Shilov boundary of D by $\partial_S D$. We shall not formally define here the notion of the Shilov boundary of a uniform algebra; we shall merely state that the Bergman–Shilov boundary of a bounded domain $D \Subset \mathbb{C}^n$ is the Shilov boundary of the uniform algebra $A(D) := \mathcal{O}(D) \cap \mathscr{C}(\overline{D})$. However, we do carefully state the following definition:

Definition 3.4. Let D be a bounded domain in $\mathbb{C}^n$. An *affine ∂D-component* is an equivalence class under the equivalence relation $\sim_A$ on ∂D given by

$$x \sim_A y \iff x \text{ and } y \text{ can be joined by a chain of segments lying in } \partial D,$$

where a segment is a subset of $\mathbb{C}^n$ of the form $\{u + tv : t \in (0, 1)\}$, $u, v \in \mathbb{C}^n$. A *holomorphic arc component of* ∂D is an equivalence class under the equivalence relation $\sim_H$ on ∂D given by

$$x \sim_H y \iff x \text{ and } y \text{ can be joined by a chain of analytic discs lying in } \partial D.$$

Roughly speaking, given a bounded domain $D \Subset \mathbb{C}^n$ and a point $x \in \partial D$, the holomorphic arc component of ∂D containing x is the largest (germ of a) complex-analytic variety lying in ∂D that contains x. The information that we require about holomorphic boundary components is:

Result 3.5 [Loos 1977, Theorem 6.3]. *Let D be the realization of a bounded symmetric domain as a bounded convex balanced domain in $\mathbb{C}^n$.*

(i) *The affine ∂D-components and the holomorphic arc components of ∂D coincide.*

(ii) *A boundary component containing a point $x \in \partial D$ is a nonempty open region in some $\mathbb{C}$-affine subspace of positive dimension passing through x* unless *x is a maximal tripotent.*

Finally, we mention the following description of the Bergman–Shilov boundary of a bounded symmetric domain:

Result 3.6 [Loos 1977, Theorem 6.5]. *Let $D \Subset \mathbb{C}^n$ be as in Result 3.5. The Bergman–Shilov boundary of D coincides with each of the following sets*:

(i) *the set of maximal tripotents of $\mathbb{C}^n$ with respect to $\{\cdot\,,\cdot\,,\cdot\}_D$*;

(ii) *the set of extreme points of $\overline{D}$*;

(iii) *the set of points of $\overline{D}$ having the maximum Euclidean distance from* $0 \in \mathbb{C}^n$.

4. Some essential propositions

This section contains several lemmas and propositions — some being simple consequences of known results, and some requiring substantial work — that will be needed to prove our theorems. We begin with the following result of Bell:

Result 4.1 [Bell 1982, Theorem 2]. *Suppose $f : D_1 \to D_2$ is a proper holomorphic map between bounded circular domains. Suppose further that D_2 contains the origin and that the Bergman kernel $K(w, z)$ associated to D_1 is such that for each compact subset G of D_1, there is an open set $U = U(G)$ containing $\overline{D}_1$ such that $K(\cdot, z)$ extends to be holomorphic on U for each $z \in G$. Then f extends holomorphically to a neighborhood of $\overline{D}_1$.*

Now let D be any bounded balanced domain (not necessarily convex) in $\mathbb{C}^n$. If D is not convex, it will not be a unit ball with respect to some norm on $\mathbb{C}^n$. However, we do have a function that has the same homogeneity property as a norm, with respect to which D is the "unit ball". The function $M_D : \mathbb{C}^n \to [0, \infty)$ defined by

$$M_D(z) := \inf\{t > 0 : z/t \in D\}$$

is called the *Minkowski functional for D*. Assume that the intersection of each complex line passing through $0 \in \mathbb{C}^n$ with ∂D is a circle. Let G be a compact subset of D. Then, as M_D is upper semicontinuous, there exists $r_G \in (0, 1)$ such that $G \subset \{z \in \mathbb{C}^n : M_D(z) < r_G\}$ and the latter is an open set. Hence $z/r_G \in D$ for all $z \in G$. Clearly, $r_G w \in D$ for all $w \in \{z \in \mathbb{C}^n : M_D(z) < 1/r_G\} =: U(G)$. By our assumptions, $\overline{D} \subset U(G)$. Let K_D be the Bergman kernel of D. We recall that

$$K_D(w, z) = \sum_{\nu \in \mathbb{N}} \psi_\nu(w)\overline{\psi_\nu(z)} \quad \text{for all } (w, z) \in D \times D,$$

where the right-hand side converges absolutely and uniformly on any compact subset of $D \times D$ and $\{\psi_\nu\}_{\nu \in \mathbb{N}}$ is *any* complete orthonormal system for the Bergman space of D. Then — owing to the fact that the collection $\{C_\alpha z^\alpha : \alpha \in \mathbb{N}^n\}$ (where $C_\alpha > 0$ are suitable normalization constants) is a complete orthonormal system for the Bergman space of D — we can infer two things. First, the functions

$$\phi_z(w) := K_D(r_G w, z/r_G), \quad w \in U(G), \tag{4-1}$$

are well-defined by power series for each $z \in G$. Secondly,

$$K_D(r_G w, z/r_G) = K_D(w, z) \quad \text{for all } (w, z) \in D \times G.$$

Comparing this with (4-1), we see that each ϕ_z extends $K_D(\cdot, z)$ holomorphically. In view of Result 4.1, we have just deduced:

Lemma 4.2. *Let $f : D_1 \to D_2$ be a proper holomorphic map between bounded circular domains. Suppose D_1 and D_2 are both balanced. Assume that the intersection of every complex line passing through 0 with ∂D_1 is a circle. Then f extends holomorphically to a neighborhood of $\overline{D}_1$.*

We remark that the above conclusion also follows from [Bell 1993].

Let D be a bounded symmetric domain in $\mathbb{C}^n$. Let $(\mathbb{C}^n, \{\cdot, \cdot, \cdot\}_D)$ be the Jordan triple system associated to D (as in other places in this paper, we assume that D

is a Harish-Chandra realization). Let D_D and Q_D be the maps (2-2) for the triple product $\{\cdot, \cdot, \cdot\}_D$. We define the linear operators $\mathsf{B}_D(x, y) : \mathbb{C}^n \to \mathbb{C}^n$, where

$$\mathsf{B}_D(x, y) := \mathsf{id}_D - \mathsf{D}_D(x, y) + Q(x)Q(y), \quad x, y \in \mathbb{C}^n.$$

Consider the sesquilinear form $(x, y) \mapsto \mathrm{Tr}[\mathsf{D}_D(x, y)]$ on $\mathbb{C}^n$. It turns out that the positivity of $\{\cdot, \cdot, \cdot\}_D$ is equivalent to this sesquilinear form being an inner product on $\mathbb{C}^n$; see [Loos 1977, Chapter 3]. Furthermore with respect to this inner product, we have

$$\mathsf{B}_D(x, y)^* = \mathsf{B}_D(y, x) \quad \text{for all } x, y \in \mathbb{C}^n.$$

It is now easy to deduce that $\mathsf{B}_D(a, a)$ is a self-adjoint, positive semidefinite linear operator. Consequently, $\mathsf{B}_D(a, a)$ admits a unique positive semidefinite square root, which we denote by $\mathsf{B}_D(a, a)^{1/2}$. Having made these two definitions, we can state the following useful facts about the geometry of D.

Result 4.3 [Loos 1977, Proposition 9.8; Roos 2000, Proposition III.4.1]. *Let D be the realization of a bounded symmetric domain as a convex balanced domain in $\mathbb{C}^n$. Fix a point $a \in D$ and let*

$$g_a(z) := a + \mathsf{B}_D(a, a)^{1/2}(\mathsf{id}_D + \mathsf{D}_D(z, a))^{-1}(z) \quad \textit{for all } z \in D.$$

Then, $g_a \in \mathrm{Aut}(D)$, $g_a(0) = a$, and $g_a'(z) = \mathsf{B}_D(a, a)^{1/2} \circ \mathsf{B}_D(z, -a)^{-1}$. Furthermore, $g_a^{-1} = g_{-a}$.

Various versions of the following lemma have been known for a long time. We refer the reader to [Rudin 1980, Lemma 15.2.2] for a proof.

Lemma 4.4. *Let D be a bounded domain in $\mathbb{C}^n$ and let $p \in \partial D$. Assume that there exists a ball B centered at p and a function $h \in \mathcal{O}(B \cap D) \cap \mathscr{C}(\overline{B \cap D}; \mathbb{C})$ such that $h(p) = 1$ and $|h(z)| < 1$ for all $z \in \overline{B \cap D} \setminus \{p\}$. Let $a_0 \in D$ and $\{\phi_k\}$ be a sequence of automorphisms of D such that $\phi_k(a_0) \to p$ as $k \to \infty$. Then, $\{\phi_k\}$ converges uniformly on compact subsets of D to const_p — the map that takes the constant value p.*

We now state a version of Schwarz's lemma for convex balanced domains and then a version of Schwarz's lemma for irreducible bounded symmetric domains, both of which are needed in the proof of our Key Lemma 1.2 (see Section 1).

Result 4.5 [Rudin 1980, Theorem 8.1.2]. *Let Ω_1 and Ω_2 be balanced regions in $\mathbb{C}^n$ and $\mathbb{C}^m$ respectively, and let $F : \Omega_1 \to \Omega_2$ be a holomorphic map. Suppose Ω_2 is convex and bounded. Then:*

(i) *$F'(0)$ maps Ω_1 into Ω_2.*

(ii) *$F(r\Omega_1) \subseteq r\Omega_2$ $(0 < r \leq 1)$ if $F(0) = 0$.*

Result 4.6 [Vigué 1991, Théorème 7.4]. *Let D be an irreducible bounded symmetric domain in $\mathbb{C}^n$ in its Harish-Chandra realization (whence it is the unit ball in $\mathbb{C}^n$ for the associated spectral norm $\|\cdot\|$). Let $F : D \to D$ be a holomorphic map such that $F(0) = 0$. Assume that for some nonempty open set $U \subset D$, we have $\|F(z)\| = \|z\|$ for all $z \in U$. Then F is an automorphism of D.*

With these two results, we can now give a proof of our Key Lemma 1.2:

The proof of the Key Lemma 1.2. Let $z \in U$, and set $w := F(z)$. By hypothesis, F maps Δ_z into D and $(F|_{W_1})^{-1}$ maps Δ_w into D. Applying Result 4.5 to $F|_{\Delta_z}$ and to $(F|_{W_1})^{-1}\big|_{\Delta_w}$, we have $\|F(z)\| = \|z\|$ for every $z \in U$. Thus by the Schwarz lemma for irreducible bounded symmetric domains, F is an automorphism of D. □

We now state and prove a technical proposition regarding the invertibility of the operator B_D at certain off-diagonal points in $\partial D \times \partial D$, where D is an *irreducible* bounded symmetric domain of dimension at least 2. Here $\mathcal{M}_{D,1}$ denotes the stratum of ∂D described by Result 3.3. This result and our Key Lemma 1.2 are the central ingredients in the proof of our main theorem.

Proposition 4.7. *Let D be the realization of an irreducible bounded symmetric domain of dimension n as a bounded convex balanced domain in $\mathbb{C}^n$, $n \geq 2$. Let $p \in \partial D$. For each $z_0 \in \mathcal{M}_{D,1}$ and each $\mathcal{M}_{D,1}$-open neighborhood $U \ni z_0$, there exists a point $w \in U$ such that* $\det \mathsf{B}_D(\cdot, p)$ *is nonzero on the set $\{\zeta w : \zeta \in \mathbb{C}, |\zeta| = 1\}$.*

Remark 4.8. In the following proof, we argue by assuming that the conclusion is false. We can instantly arrive at a contradiction at the point (•) in the proof below if we invoke results on the fine structure of ∂D; see [Wolf and Korányi 1965] or [Wolf 1972], for instance. However, we provide an elementary argument beyond (•) to complete the proof in the hope that appropriate analogues of Proposition 4.7 may be formulated in other contexts.

Proof. Let us denote $\det \mathsf{B}_D(z, p)$ by $h(z)$, where $z \in \mathbb{C}^n$. Let us assume that the result is false. Then, there exist a point $z_0 \in \mathcal{M}_{D,1}$ and an $\mathcal{M}_{D,1}$-open neighborhood $U \ni z_0$ such that for each $w \in U$, there exists a $\zeta_w \in \{\zeta \in \mathbb{C} : |\zeta| = 1\}$ with $h(\zeta_w w) = 0$. Let q denote the quotient map $q : \mathbb{C}^n \setminus \{0\} \to \mathbb{CP}^{n-1}$. Also write

$$Z_h := h^{-1}\{0\}, \quad Z := Z_h \cap \mathcal{M}_{D,1}.$$

Our assumption implies that $q(Z)$ contains a nonempty open set $\mathcal{V} \subset \mathbb{CP}^{n-1}$. Let $\mathcal{A} := \{z \in \mathbb{C}^n : 1 - \varepsilon < \|z\| < 1 + \varepsilon\}$, where $\|\cdot\|$ denotes the spectral norm relative to which D is the unit ball and ε is a *fixed* number in $(0, 1)$. As $\mathcal{V} \subset q(\mathcal{A})$, it is easy to see that $\mathcal{V}$ can be covered by finitely many holomorphic coordinate patches $(U_1, \psi_1), \ldots, (U_M, \psi_M)$ such that the maps

$$q_j := \psi_j \circ q|_{q^{-1}(U_j) \cap \mathcal{A}} : q^{-1}(U_j) \cap \mathcal{A} \to \mathbb{C}^{n-1}$$

are Lipschitz maps. Since Lipschitz maps cannot increase Hausdorff dimension (see [Rudin 1980, Proposition 14.4.4], for instance) and $\dim_{\mathbb{R}}(\mathscr{V}) = 2n-2$, the preceding discussion shows that the Hausdorff dimension of Z (and hence the dimension of Z as a real-analytic set) is $2n-2$. As Z_h is a complex analytic subvariety, its singular locus is of complex dimension at most $n-2$. Thus, we can find a point $x_0 \in Z$ that is a regular point of Z_h and an open ball B around x_0 that is so small that:

- $\mathcal{M}_{D,1} \cap B$ is a submanifold of B;
- $B \cap Z_h$ is an $(n-1)$-dimensional complex submanifold of B;
- the dimension of $B \cap Z$ is $2n-2$.

These three facts imply that $M := B \cap Z_h \subset \mathcal{M}_{D,1}$. We can deduce this by considering a local defining function $\rho_B : B \to \mathbb{R}$ for $\mathcal{M}_{D,1}$ and observing that, by Łojasiewicz's theorem [1959], $\rho_B|_M \equiv 0$. *If $D = \mathbb{B}^n$, we already have a contradiction* and, hence, the proof.

Since $\mathcal{M}_{D,1}$ is a real-analytic submanifold of $\mathbb{C}^n \setminus \bigsqcup_{j \geq 2} \mathcal{M}_{D,\,j}$, where $\mathcal{M}_{D,\,j}$ are the strata of ∂D discussed in Section 3, we can define the Levi-form of $\mathcal{M}_{D,1}$ — denoted by $\mathfrak{L}(z, V)$, where $z \in \mathcal{M}_{D,1}$, $V \in H_z(\mathcal{M}_{D,1})$. A few words about notation: in this proof, we shall work with the tangent bundle of $\mathcal{M}_{D,1}$ defined extrinsically. So, when referring to vectors in $T_z(\mathcal{M}_{D,1})$, we shall view them either as real or as complex vectors, as convenient, such that $z + T_z(\mathcal{M}_{D,1})$ is the hyperplane tangent to $\mathcal{M}_{D,1}$ at $z \in \mathcal{M}_{D,1}$. In this scheme,

$$H_z(\mathcal{M}_{D,1}) := T_z(\mathcal{M}_{D,1}) \cap i T_z(\mathcal{M}_{D,1}).$$

As $\dim_{\mathbb{C}}(M) = n-1$, we have $\mathfrak{L}(z, \cdot\,) \equiv 0$ for all $z \in M$. The curve $\gamma : (-\varepsilon, \varepsilon) \to \mathcal{M}_{D,1}$ (for $\varepsilon > 0$ suitably small), $\gamma(t) := \exp(it)z$, is transverse to M at z. This is because if $\gamma'(0) = iz$ were in $H_z(\mathcal{M}_{D,1})$, then

$$i\gamma'(0) = -z \in H_z(\mathcal{M}_{D,1}) \subset T_z(\mathcal{M}_{D,1}),$$

which contradicts the convexity of D. Consequently, for $\varepsilon_0 > 0$ sufficiently small, the set $\{\exp(it)z : t \in (-\varepsilon_0, \varepsilon_0), z \in M\}$ contains an $\mathcal{M}_{D,1}$-open neighborhood of x_0. Thus, $\mathcal{M}_{D,1}$ is Levi-flat at x_0. As $\mathcal{M}_{D,1}$ is real-analytic, it is a Levi-flat hypersurface.

We shall now show that the Levi-flatness of $\mathcal{M}_{D,1}$ leads to a contradiction. Let us pick an $x \in \mathcal{M}_{D,1}$. Owing to Levi-flatness, we can find a ball B_x, centered at x, such that

$$D_x^- := D \cap B_x, \quad D_x^+ := B_x \setminus \overline{D}$$

are both pseudoconvex. Let $\boldsymbol{n}_x$ denote the unit outward normal vector to ∂D at x ($x \in \mathcal{M}_{D,1}$). Owing to convexity of D, we can find an $\varepsilon_0 > 0$ and a $\delta_0 > 0$ such that

$$H_x(\varepsilon_0; \delta) := x + \delta\boldsymbol{n}_x + \{V \in H_x(\mathcal{M}_{D,1}) : |V| < \varepsilon_0\} \subset D_x^+$$

for each $\delta \in (0, \delta_0)$. Here, $|\cdot|$ denotes the Euclidean norm. As $H_x(\varepsilon_0; \delta)$ is a copy of a complex $(n-1)$-dimensional ball and as D_x^+ is taut — see [Kerzman and Rosay 1981, Proposition 2.1] — it follows that $H_x(\varepsilon_0; 0) \subset \mathcal{M}_{D,1}$. To summarize, $\mathcal{M}_{D,1}$ has the following property:

(•) At each $x \in \mathcal{M}_{D,1}$, a germ of the set $(x + H_x(\mathcal{M}_{D,1}))$ lies in $\mathcal{M}_{D,1}$.

Let us now pick and fix a point $y^0 \in \mathcal{M}_{D,1}$. Let $(z_1, \dots, z_n)$ be global holomorphic coordinates in $\mathbb{C}^n$, associated to an appropriate rigid motion of D, such that $y^0 = (0, \dots, 0)$, $D \subset \{\mathrm{Re}\, z_1 > 0\}$ and $H_{y^0}(\mathcal{M}_{D,1}) = \{z_1 = 0\}$ relative to these coordinates. Let W be a nonzero vector in $H_{y^0}(\mathcal{M}_{D,1})$ and let $D_W := D \cap \mathrm{span}_{\mathbb{C}}\{W, \boldsymbol{n}_{y^0}\}$. Clearly, D_W is convex and by (•), $\mathcal{M}_{D,1} \cap \mathrm{span}_{\mathbb{C}}\{W, \boldsymbol{n}_{y^0}\} =: \mathcal{M}_W$ has the property that for each point $y \in \mathcal{M}_W$, the germ of a complex line through y, call it $\Lambda_{y,W}$, lies in $\mathcal{M}_W$. Let us view D_W as lying in $\mathbb{C}^2$, whence a portion of $\mathcal{M}_W$ near $(0, 0)$ can be parametrized by three real variables as follows:

$$r(t, u, v) = \rho(t) + a(t)(u + iv), \qquad |t| < \varepsilon_1,\ \ |u|, |v| < \varepsilon_2,$$

where $\rho : (-\varepsilon_1, \varepsilon_1) \to \mathcal{M}_W$ is a smooth curve through $(0, 0)$ such that $\rho'(t)$ is orthogonal to $\Lambda_{\rho(t),W}$ for each t, and $a : (-\varepsilon_1, \varepsilon_1) \to \mathbb{C}^2$ is such that $a(t)$ is parallel to $\Lambda_{\rho(t),W}$ for each t. For the remainder of this paragraph, $\boldsymbol{n}(t, u, v)$ will denote the *inward* unit normal to ∂D_W at $r(t, u, v)$, and $\cdot$ will denote the standard inner product on $\mathbb{R}^4$. Define the matrix-valued function $\Gamma : (-\varepsilon_1, \varepsilon_1) \times (-\varepsilon_2, \varepsilon_2)^2 \to \mathbb{R}^{3\times 3}$ by

$$\Gamma(\tau, U, V) := \mathsf{Hess}_{t,u,v}(r(t, u, v) \cdot \boldsymbol{n}(\tau, U, V))\big|_{(t,u,v)=(\tau,U,V)}\ .$$

The convexity of D_W implies that $\Gamma(\tau, U, V)$ is *positive* semidefinite at each (τ, U, V) (recall that $\boldsymbol{n}(\tau, U, V)$ is the inward normal at $r(\tau, U, V)$). By choosing $\varepsilon_1, \varepsilon_2 > 0$ small enough, we can ensure that $(n_1^2 + n_2^2)(t, u, v) \neq 0$ for every (t, u, v), where we write $\boldsymbol{n} = (n_1, n_2, n_3, n_4)$, and that a is of the form $a(t) = (\alpha(t) + i\beta(t), 1)$. We compute to observe that two of the principal minors of Γ turn out to be $-(n_1\alpha' + n_2\beta')^2$ and $-(n_2\alpha' - n_1\beta')^2$, which must be nonnegative. This gives us the system of equations

$$\begin{aligned} n_1\alpha' + n_2\beta'\big|_{(\tau,U,V)} &= 0 \\ -n_1\beta' + n_2\alpha'\big|_{(\tau,U,V)} &= 0 \quad \text{for all } (\tau, U, V). \end{aligned}$$

By our assumption on $\boldsymbol{n}$, this implies that $\alpha' = \beta' \equiv 0$. Restating this geometrically, there is a small $\mathcal{M}_W$-open neighborhood of $0 \in \partial D_W$ such that, for every y in this neighborhood, $\Lambda_{y,W}$ is parallel to the vector W. This holds true for each nonzero $W \in H_{y^0}(\mathcal{M}_{D,1})$. Thus, there is an $\mathcal{M}_{D,1}$-open patch $\omega \ni y^0$ such that

$$x + H_x(\mathcal{M}_{D,1}) \text{ is parallel to } \{z_1 = 0\} \text{ for every } x \in \omega. \tag{4-2}$$

By Result 3.3, $\mathcal{M}_{D,1}$ is connected. Thus, if $y^0 \neq y \in \mathcal{M}_{D,1}$, then y can be joined to y^0 by a chain of $\mathcal{M}_{D,1}$-open patches $\omega_0, \dots, \omega_N$, where ω_0 equals the patch ω in (4-2), $\omega_{j-1} j \cap \omega_j \neq \varnothing$, $j = 1, \dots, N$, and $\omega_N \ni y$. By a standard argument of real-analytic continuation, we deduce that (4-2) holds with ω_N replacing ω (where z_1 comes from the global system of coordinates fixed at the beginning of the previous paragraph). Hence, $x + H_x(\mathcal{M}_{D,1})$ is parallel to $\{z_1 = 0\}$ for each $x \in \mathcal{M}_{D,1}$. As $\mathcal{M}_{D,1}$ is dense in ∂D and D is bounded, we can find a $\xi \in D$ and a vector $W = (W_1, \dots, W_n)$ with $W_1 = 0$ such that the ray $\{\xi + tW : t \geq 0\}$ intersects ∂D at a point in $\mathcal{M}_{D,1}$. Then, this ray must be tangential to $\mathcal{M}_{D,1}$ at the point of intersection, which is absurd as D is convex. Hence, our initial assumption must be false. $\square$

5. The proof of Theorem 1.1

Before we proceed further, we clarify our notation for the different norms that will be used in the proof of Theorem 1.1. With D_1 and D_2 as in Theorem 1.1, $\|\cdot\|_j$ will denote the spectral norm such that D_j is the unit $\|\cdot\|_j$-ball in $\mathbb{C}^n$, $j = 1, 2$. The Euclidean norm on $\mathbb{C}^n$ will be denoted by $|\cdot|$. We will also need to impose norms on certain linear operators on $\mathbb{C}^n$. We shall use the operator norm induced by the Euclidean norm: for a $\mathbb{C}$-linear operator A on $\mathbb{C}^n$, we set

$$\|A\|_{\mathrm{op}} := \sup_{|x|=1} |Ax|.$$

Proof of Theorem 1.1. We shall take D_1 and D_2 to be Harish-Chandra realizations of the given bounded symmetric domains. We may assume, composing F with suitable automorphisms if necessary, that $F(0) = 0$.

By Lemma 4.2, F extends to a holomorphic map defined on a neighborhood N of $\overline{D}_1$. For simplicity of notation, we shall denote this extension also by F. The complex Jacobian $\mathrm{Jac}_{\mathbb{C}} F$ is holomorphic on N and $\mathrm{Jac}_{\mathbb{C}} F \not\equiv 0$ on D_1. Hence, by the maximum principle, $\mathrm{Jac}_{\mathbb{C}} F \not\equiv 0$ on ∂D_1. By definition, we can find a point $p \in \partial_S D_1$ such that

$$\sup_{\overline{D}_1} |\mathrm{Jac}_{\mathbb{C}} F| = |\mathrm{Jac}_{\mathbb{C}} F(p)| \neq 0.$$

By the inverse function theorem, we can find a ball $B(p, r) \subset N$ such that $F|_{B(p,r)}$ is injective. Let us write

$$\Omega_1 := B(p, r) \cap D_1, \quad \Omega_2 := F(B(p, r)) \cap D_2.$$

We shall use our Key Lemma 1.2 (see Section 1 and its proof in Section 4) to deduce the result. The regions W_1 and W_2 of that lemma will be constructed by applying suitable automorphisms to Ω_1 and Ω_2.

Claim. $$F(p) \in \partial_S D_2.$$

Suppose $F(p) \notin \partial_S D_2$. It follows from Result 3.5 and Result 3.6 that there are a vector $V \in \mathbb{C}^n \setminus \{0\}$ and neighborhood ω of $0 \in \mathbb{C}$ such that $\psi(\omega) \subset F(B(p,r)) \cap \partial D_2$, where $\psi : \omega \ni \zeta \mapsto F(p) + \zeta V$. Next, define

$$\widetilde{\psi} := (F|_{B(p,r)})^{-1} \circ \psi.$$

Since $F|_{D_1}$ is proper and $F|_{B(p,r)}$ is injective,

$$F(z) \in F(B(p,r)) \cap \partial D_2 \iff z \in B(p,r) \cap \partial D_1.$$

Thus $\widetilde{\psi}(\omega) \subset \partial D_1$. Furthermore, $\widetilde{\psi}$ is nonconstant and $\widetilde{\psi}(0) = p$. By definition, each point of $\widetilde{\psi}(\omega) \setminus \{p\}$ lies in the holomorphic arc component of ∂D_1 containing p. This is a contradiction since p, being an extreme point, is a one-point affine ∂D_1-component and thus, by Result 3.5, a one-point holomorphic arc component of ∂D_1. Hence the claim.

Let us now take a sequence $\{a_k\} \subset \Omega_1$ such that $a_k \to p$, and let $b_k := F(a_k)$. Let $\phi_k^1 \in \mathrm{Aut}(D_1)$ denote an automorphism that maps 0 to a_k. Let $\phi_k^2 \in \mathrm{Aut}(D_2)$ be an automorphism that maps 0 to b_k. Owing to Result 3.6 and to convexity, we can construct a peak function for p on $\overline{D}_1$. Likewise (in view of the last claim) $F(p)$ is a peak point of D_2. By Lemma 4.4, we get

$$\phi_k^j \to \mathsf{const}_{p^j} \ \text{ uniformly on compacts}, \quad j = 1, 2, \tag{5-1}$$

where $p^1 := p$ and $p^2 := F(p)$.

We now define

$$\Omega_j^k := (\phi_k^j)^{-1}(\Omega_j), \quad j = 1, 2, \ k \in \mathbb{Z}_+.$$

Given any $r > 0$, write $rD_j := \{z \in \mathbb{C}^n : \|z\|_j < r\}$, $j = 1, 2$. By (5-1), there exists a sequence $k_1 < k_2 < k_3 < \cdots$ in $\mathbb{Z}_+$ such that

$$\phi_{k_\nu}^1((1 - 1/s)\overline{D}_1) \subset \Omega_1 \quad \text{for all } \nu \geq s, \ s \in \mathbb{Z}_+.$$

By (5-1) again, we can extract a sequence of indices $\nu(1) < \nu(2) < \nu(3) < \cdots$ such that

$$\phi_{k_{\nu(t)}}^2((1 - 1/s)\overline{D}_2) \subset \Omega_2 \quad \text{for all } t \geq s, \ s \in \mathbb{Z}_+.$$

In the interest of readability of notation, let us reindex $\{k_{\nu(s)}\}_{s \in \mathbb{Z}_+}$ as $\{k_m\}_{m \in \mathbb{Z}_+}$. Then, the above can be summarized as

(∗) With the sequences of maps $\{\phi_k^1\} \subset \mathrm{Aut}(D_1)$ and $\{\phi_k^2\} \subset \mathrm{Aut}(D_2)$ as described above, there are a sequence $\{k_m\}_{m \in \mathbb{Z}_+} \subset \mathbb{Z}_+$ and a strictly increasing $\mathbb{Z}_+$-valued function ν^* such that

$$(1 - 1/s)\overline{D}_1 \subset \Omega_1^{k_m} \quad \text{for all } m \geq s, \ s \in \mathbb{Z}_+,$$
$$(1 - 1/\nu^*(s))\overline{D}_2 \subset \Omega_2^{k_m} \quad \text{for all } m \geq s, \ s \in \mathbb{Z}_+.$$

Step 1. *Analyzing the family* $\{(\phi^2_{k_m})^{-1} \circ F \circ \phi^1_{k_m}\}_{m \in \mathbb{Z}_+}$.
Consider the maps $G_m : D_1 \to D_2$ defined by

$$G_m := (\phi^2_{k_m})^{-1} \circ F \circ \phi^1_{k_m}.$$

By Montel's theorem, and passing to a subsequence and relabeling if necessary, we get a map $G \in \mathcal{O}(D_1; \mathbb{C}^n)$ such that $G_m \to G$ uniformly on compact subsets. Let us fix an $s \in \mathbb{Z}_+$. By $(*)$, we infer that there exists $M_s \in \mathbb{Z}_+$ such that $(1-1/s)\overline{D}_j \subset \Omega^{k_m}_j$ for all $m \geq M_s$, $j = 1, 2$. Note that $G_m|_{\Omega^{k_m}_1}$ is a biholomorphism, whence $G'_m(0)$ is invertible for each m. Hence, by the Schwarz lemma for convex balanced domains (i.e., Result 4.5 above) $G'_m(0)$ maps $(1-1/s)D_1$ into D_2 and $G'_m(0)^{-1}$ maps $(1-1/s)D_2$ into D_1 for all $m \geq M_s$. We claim that this implies that $G'(0)$ is invertible. Suppose not. Then we would find a z_0 with $\|z_0\|_1 = (1-2/s)$ such that $G'(0)z_0 = 0$. Note that $G'_m(0) \to G'(0)$ in norm, whence, given any $\varepsilon > 0$, $\|G'_m(0)z_0\|_2 < \varepsilon$ for every sufficiently large m. If we now choose $\varepsilon \leq (1-2/s)^2$, we see that

$$G'_m(0)^{-1}\big(\{\|w\|_2 = (1-2/s)\}\big) \not\subset D_1$$

for all sufficiently large m. This is a contradiction. Hence the claim.

Now that it is established that $G'(0)$ is invertible, it follows that $G'_m(0)^{-1} \to G'(0)^{-1}$ in norm. Hence, $G'(0)^{-1}$ maps $(1-1/s)D_2$ into D_1. Recall that $s \in \mathbb{Z}_+$ was arbitrarily chosen and that the function ν^* in $(*)$ is strictly increasing. Thus, $G'(0)^{-1}$ maps D_2 into D_1. By construction, $G(D_1) \subset \overline{D}_2$. Now, D_2 is complete (Kobayashi) hyperbolic. Hence D_2 is taut; see [Kiernan 1970]. As $G(0) = 0 \in D_2$, G maps D_1 to D_2. So, the holomorphic map $G'(0)^{-1} \circ G : D_1 \to D_1$ satisfies all the conditions of Cartan's uniqueness theorem. Thus,

$$G'(0)^{-1} \circ G = \mathsf{id}_{D_1},$$

which means that $G = G'(0)|_{D_1}$.

Step 2. *Showing that D_1 and D_2 are biholomorphically equivalent.*
We have shown in Step 1 that $G'(0)^{-1}$ maps $(1-1/s)D_2$ into D_1. As $G'(0)$ is injective, this means that $G'(0)(D_1)$ contains $(1-1/s)D_2$ for arbitrarily large $s \in \mathbb{Z}_+$. Thus G maps D_1 onto D_2. It follows that D_1 is biholomorphic to D_2.

It would help to simplify our notation somewhat. By the nature of the argument in Step 1, it is clear that we can assume that the sequences $\{a_k\} \subset \Omega_1$ and $\{b_k\} \subset \Omega_2$ are so selected that $(*)$ is true with $\{k_m\}_{m \in \mathbb{Z}_+} = \{1, 2, 3, \dots\}$. Owing to Step 2, we may now assume $D_1 = D_2 =: D$. The argument we will make in Step 3 below is valid regardless of the specific sequence $\{a_k\}$ or $\{b_k\}$. Hence, in the next three paragraphs following this, the sequence $\{A_k\}$ will stand for either $\{a_k\}$ or $\{b_k\}$, and the point q will stand for either p or $F(p)$. Also, we will abbreviate $\phi^j_{A_k}$ to ϕ_k.

Step 3. *Producing subsequences of* $\{\phi_k\}$ *that converge on "large" subsets of* ∂D. By Result 4.3 we may take $\phi_k = g_{A_k}$, whence

$$\phi_k'(z) = \mathsf{B}_D(A_k, A_k)^{1/2} \circ \mathsf{B}_D(-z, A_k)^{-1}. \tag{5-2}$$

In the argument that follows, it is implicit that each ϕ_k is defined as a holomorphic map on some neighborhood (which depends on ϕ_k) of $\overline{D}$; see Lemma 4.2. By Proposition 4.7 we can find a point $\xi_0 \in \mathcal{M}_{D,1}$ such that

$$\det \mathsf{B}_D(e^{i\theta}\xi_0, q) \neq 0 \quad \text{for all } \theta \in \mathbb{R}.$$

By continuity, there exist a $\overline{D}$-open neighborhood Γ of q, an $\mathcal{M}_{D,1}$-open neighborhood W of ξ_0, and a $\overline{D}$-open set V with the following properties:

(a) $z \in V \implies e^{i\theta}z \in V$ for all $\theta \in \mathbb{R}$,

(b) $V \cap \partial D = S^1 \cdot W$,

(c) $z \in V \implies tz \in V$ for all $t \in [1, 1/\|z\|]$

(now $\|\cdot\|$ is the spectral norm associated to D), such that

$$\det \mathsf{B}_D(z, w) \neq 0 \quad \text{for all } (z, w) \in \overline{V} \times \Gamma. \tag{5-3}$$

Here, given a set $X \subset \mathbb{C}^n$, $S^1 \cdot X$ stands for the set $\{e^{i\theta}x : x \in X, \theta \in \mathbb{R}\}$. Let us call any pair (V, W), where V is a $\overline{D}$-open set and W is an $\mathcal{M}_{D,1}$-open set, a *truncated prism with base* $S^1 \cdot W$ if (V, W) satisfies properties (a)–(c) above.

We can find V' and W', with $\overline{W'} \subset W$, such that (V', W') is a truncated prism with base $S^1 \cdot W'$ with the properties

- $\overline{V'} \subset V$;
- there exists a $\delta_0 \ll 1$ such that for $z_1, z_2 \in \overline{V'}$, the segment $[z_1, z_2] \subset V$ whenever $|z_1 - z_2| < \delta_0$.

Owing to holomorphicity and convexity,

$$\phi_k(z_1) - \phi_k(z_2) = \int_0^1 \phi_k'(z_1 + t(z_2 - z_1))(z_2 - z_1)\, dt, \quad z_1, z_2 \in \overline{D}. \tag{5-4}$$

We can find a $K \equiv K(W)$ such that, in view of (5-3), $\{\mathsf{B}_D(z, A_k) : k \geq K(W), z \in \overline{V}\}$ is a compact family in $\mathrm{GL}(n, \mathbb{C})$. Hence, in view of (5-2) (and since $\{\mathsf{B}_D(A_k, A_k) : k \in \mathbb{Z}_+\}$ is a relatively compact family in $\mathbb{C}^{n\times n}$), there exists a constant $C > 0$ such that

$$\|\phi_k'(z)\|_{\mathrm{op}} \leq C \quad \text{for all } z \in \overline{V} \text{ and } k \geq K.$$

By our construction of V' and from (5-4), we conclude that

$$|\phi_k(z_1) - \phi_k(z_2)| \leq C|z_1 - z_2| \quad \text{for all } z_1, z_2 \in \overline{V'}, |z_1 - z_2| < \delta_0, \text{ and } k \geq K.$$

In short, $\{\phi_k|_{\overline{V'}}\} \subset \mathscr{C}(\overline{V'}; \mathbb{C}^n)$ is an equicontinuous family.

By the Arzelà–Ascoli theorem, we can find a subsequence of $\{\phi_k\}$ that converges uniformly to q on $\overline{V}'$. For simplicity of notation, let us continue to denote this subsequence by $\{\phi_k\}$. Then there exists a $K_1 \in \mathbb{Z}_+$ such that $\phi_k(\overline{V}') \subset \Omega$ (which denotes either Ω_1 or Ω_2) for all $k \geq K_1$. Furthermore, we may assume that K_1 is so large that, thanks to $(*)$,

$$(1-1/s)\overline{D} \subset \phi_k^{-1}(\Omega) \quad \text{for all } k \geq K_1,$$

where s is so large that $(1-1/s)\overline{D} \cap V'$ is a nonempty open set. By construction,

$$z \in V' \cap D \implies \Delta_z \subset (1-1/s)\overline{D} \cup V'.$$

Hence $\Delta_z \subset \phi_k^{-1}(\Omega)$ for all $k \geq K_1$. We summarize this paragraph as follows:

$(**)$ Given any truncated prism (V, W) with base $S^1 \cdot W$ such that $\mathsf{B}_D(z, A_k) \neq 0$ on $\overline{V}$ for all k sufficiently large, we can find a $K_1 \in \mathbb{Z}_+$ and a truncated prism (V', W') with $\overline{V}' \subset V$ such that $\Delta_z \subset \phi_k^{-1}(\Omega)$ for each $z \in V' \cap D$ and each $k \geq K_1$.

Step 4. *Completing the proof.*
By Proposition 4.7 and $(**)$, we can find a truncated prism (V', W') with base $S^1 \cdot W'$ which has all the properties stated in $(**)$. Let $s \in \mathbb{Z}_+$ be so large that $(1-1/s)D \cap V' := U'$ is a nonempty open set. As $G_k \to G$ uniformly on U' (by Step 1), there exist a point $w_0 \in G(U')$, $K_2 \in \mathbb{Z}_+$ and a $c > 0$ such that

$$B(w_0, c) \subset G(U') \cap G_k(U') \quad \text{and} \quad B(w_0, c) \subset \Omega_2^k \quad \text{for all } k \geq K_2.$$

Write $\|\cdot\|$ for the spectral norm associated to D. Let $R : \mathbb{C}^n \setminus 0 \to \partial D$ be given by $R(w) := w/\|w\|$. By Proposition 4.7 and $(**)$, we can find an $\mathcal{M}_{D,1}$-open subset ω_2 such that

$$\omega_2 \subset R(B(w_0, c)),$$

a truncated prism (V_2, ω_2) with base $S^1 \cdot \omega_2$, and a $K_3 \in \mathbb{Z}_+$ such that $\Delta_w \subset \Omega_2^k$ for each $w \in V_2 \cap D$ and each $k \geq K_3$. Let us now set $U := G^{-1}(R^{-1}(\omega_2) \cap B(w_0, c))$, and $K^* := \max(K_1, K_2, K_3)$. Finally, we set

$$W_j := (\phi_{K^*}^j)^{-1}(\Omega_j), \quad j = 1, 2,$$

with the understanding that $\phi_k^1 = g_{a_k}$ and $\phi_k^2 = g_{b_k}$.

As $U \subset V'$, we have $\Delta_z \subset W_1$ for each $z \in U$. By construction,

$$G_{K^*}(z) \in B(w_0, c) \subset W_2 \quad \text{for all } z \in U.$$

Finally, by construction, for each $z \in U$, there exists a point $w_z \in \Delta_{G_{K^*}(z)}$ that belongs to $V_2 \cap D$. Thus, $\Delta_{G_{K^*}(z)} \subset W_2$. Recall that $G_{K^*}|_{W_1} : W_1 \to W_2$ is a biholomorphism and $G_{K^*}(0) = 0$. By our Key Lemma 1.2, G_{K^*}, and consequently F, must be a biholomorphism. $\square$

6. The proof of Theorem 1.5

As p is an orbit accumulation point, there is a point $a_0 \in D_1$ and a sequence $\{\phi_k\} \subset \mathrm{Aut}(D_1)$ such that $\phi_k(a_0) \to p$. Regardless of whether p is a peak point or $F(p)$ is a peak point, let us denote the relevant peak function as H. Let B denote a small ball centered at p, with $B \Subset U$, if p is a peak point, and centered at $F(p)$, with $B \Subset F(U)$, if $F(p)$ is a peak point. Depending on whether p or $F(p)$ is a peak point, set $G := F^{-1}$ or $G := F$, respectively. Finally, set

$$h := \begin{cases} H \circ G|_{\overline{B \cap D_2}} & \text{if } p \text{ is a peak point,} \\ H \circ G|_{\overline{B \cap D_1}} & \text{if } F(p) \text{ is a peak point.} \end{cases}$$

By our hypothesis on F, it follows that h satisfies all the conditions required of the function h in Lemma 4.4 for the appropriate choice of (D, p) depending on whether p or $F(p)$ is a peak point.

Let us now denote the automorphisms discussed above by ϕ_k^1, $k = 1, 2, 3, \dots$. Then, using H or the function h constructed above, depending on whether p or $F(p)$ is a peak point, we deduce by Lemma 4.4 that $\phi_k^1 \to \mathsf{const}_p$ uniformly on compact subsets of D_1. Set $a_k := \phi_k^1(0)$. As $a_k \to p$, we may assume without loss of generality that $a_k \in U$. Let $b_k := F(a_k)$, and let $\phi_k^2 \in \mathrm{Aut}(D_2)$ be an automorphism that maps 0 to b_k (which is possible as $\mathrm{Aut}(D_2)$ acts transitively on D_2). Repeating the above argument, $\phi_k^2 \to \mathsf{const}_{F(p)}$ uniformly on compact subsets of D_2. We have arrived at the same result as in (5-1). Thereafter, if we define

$$\Omega_j^k := (\phi_k^j)^{-1}(\Omega_j), \quad j = 1, 2, \ k \in \mathbb{Z}_+,$$

where $\Omega_1 := U$ and $\Omega_2 := F(U)$, then, reasoning exactly as in the passage following (5-1), we deduce that $(*)$ from Section 5 holds true for our present setup.

With $\{k_m\}_{m \in \mathbb{Z}_+}$ as given by $(*)$, let us define the maps $G_m : \Omega_1^{k_m} \to \Omega_2^{k_m}$ by

$$G_m := (\phi_{k_m}^2)^{-1} \circ F \circ \phi_{k_m}^1.$$

By construction, each G_m is a biholomorphic map. In particular,

$$G_m(0) = 0 \quad \text{and} \quad G_m'(0) \in \mathrm{GL}(n, \mathbb{C}). \tag{6-1}$$

We may assume, owing to (5-1), that *the sequences* $\{\Omega_j^{k_m}\}_{m \in \mathbb{Z}_+}$ *are increasing sequences*. By Montel's theorem, and arguing by induction, we can find sequences $\{G_{l,m}\}$ and holomorphic maps $\Gamma_l : \Omega_1^{k_l} \to \overline{D}_2$ such that for $l = 1, 2, 3, \dots,$

- $\{G_{1,m}\}_{m \in \mathbb{Z}_+}$ is a subsequence of $\{G_\nu\}_{\nu \in \mathbb{Z}_+}$ and $\{G_{l+1,m}\}_{m \in \mathbb{Z}_+}$ is a subsequence of $\{G_{l,\nu}\}_{\nu \in \mathbb{Z}_+}$;
- $G_{l,m}\big|_{\Omega_1^{k_l}} \to \Gamma_l$ as $m \to \infty$, uniformly on compact subsets of $\Omega_1^{k_l}$.

Owing to this construction, the rule

$$\Gamma(z) := \Gamma_l(z) \quad \text{ if } z \in \Omega_1^{k_l}$$

gives a well-defined holomorphic map $\Gamma : D_1 \to \overline{D}_2$.

Let us define $H_l := G_{l,l}$. Now suppose $\Gamma(D_1) \cap \partial D_2 \neq \varnothing$. Then, there exists $\xi \in D_1$ such that $\Gamma(\xi) \in \partial D_2$. Let $M \in \mathbb{Z}_+$ be so large that $\Omega_1^{k_M} \ni \xi$. As D_2 is a bounded symmetric domain, it is taut. Thus, by focusing attention on the sequence

$$\{ H_l|_{\Omega_1^{k_M}} : l = M, M+1, M+2, \dots \} \subset \mathcal{O}(\Omega_1^{k_M}; D_2),$$

we conclude, by assumption, that $\Gamma(\Omega_1^{k_M}) \subset \partial D_2$. But, by (6-1), $\Gamma(0) = 0 \notin \partial D_2$. This is a contradiction, from which we infer:

(a) The range of Γ is a subset of D_2.

Now observe that, by $(*)$, we have:

(b) The sequence $\{H_l : l = s, s+1, s+2, \dots\}$ converges uniformly to Γ on $(1-1/s)\overline{D}_1$, $s \in \mathbb{Z}_+$;

(c) H_l^{-1} maps 0 to 0 and $(1-1/\nu^*(l))D_2$ into D_1 (since $\mathsf{dom}(H_l^{-1}) = \mathsf{range}(H_l) \supseteq \Omega_2^{k_l}$).

In view of (6-1) and the fact that D_1 and D_2 are balanced, (a)–(c) are precisely the ingredients required to repeat the argument in Step 1 of the proof of Theorem 1.1 to infer that $\Gamma'(0)$ is invertible, $\Gamma'(0)^{-1} : D_2 \to D_1$ and

$$\Gamma'(0)^{-1} \circ \Gamma = \mathsf{id}_{D_1}.$$

Thus, by (a), $\Gamma'(0)(D_1) \subset D_2$. One of the consequences of repeating the argument contained in Step 1 in Theorem 1.1 is, in view of (c), that $\Gamma'(0)^{-1}$ maps $(1-1/\nu^*(l))D_2$ into D_2 for every $l \in \mathbb{Z}_+$. As ν^* is strictly increasing and $\mathbb{Z}_+$-valued, and as $\Gamma'(0)$ is injective, this means that $\Gamma'(0)(D_1)$ contains $(1-1/s)D_2$ for *arbitrarily large* $s \in \mathbb{Z}_+$, whence $\Gamma'(0)$ maps D_1 onto D_2. Hence, $\Gamma'(0)|_{D_1}$ is a biholomorphism of D_1 onto D_2. □

Acknowledgements

Jaikrishnan Janardhanan would like to thank his colleagues and friends G.P. Balakumar, Dheeraj Kulkarni, Divakaran Divakaran and Pranav Haridas for many interesting discussions. Gautam Bharali contributed to this work while on sabbatical at the Norwegian University of Science and Technology (NTNU) in Trondheim. He would like to acknowledge the support and the hospitality of the Department of Mathematical Sciences at NTNU. He also thanks John Erik Fornæss for his helpful comments on this work.

References

[Alexander 1977] H. Alexander, "Proper holomorphic mappings in C^n", *Indiana Univ. Math. J.* **26**:1 (1977), 137–146. MR 54 #10685 Zbl 0391.32015

[Bell 1982] S. R. Bell, "Proper holomorphic mappings between circular domains", *Comment. Math. Helv.* **57**:4 (1982), 532–538. MR 84m:32032 Zbl 0511.32013

[Bell 1993] S. Bell, "Algebraic mappings of circular domains in $\mathbb{C}^n$", pp. 126–135 in *Several complex variables* (Stockholm, 1987/1988), edited by J. E. Fornæss, Math. Notes **38**, Princeton Univ. Press, 1993. MR 94a:32040 Zbl 0774.32001

[Chern and Moser 1974] S. S. Chern and J. K. Moser, "Real hypersurfaces in complex manifolds", *Acta Math.* **133** (1974), 219–271. MR 54 #13112 Zbl 0302.32015

[Harish-Chandra 1956] Harish-Chandra, "Representations of semisimple Lie groups, VI: Integrable and square-integrable representations", *Amer. J. Math.* **78** (1956), 564–628. MR 18,490d Zbl 0072.01702

[Henkin and Novikov 1984] G. M. Henkin and R. G. Novikov, "Proper mappings of classical domain", pp. 625–627 in *Linear and complex analysis problem book: 199 research problems*, edited by V. P. Havin et al., Lecture Notes in Mathematics **1043**, Springer, Berlin, 1984. MR 85k:46001 Zbl 0545.30038

[Kerzman and Rosay 1981] N. Kerzman and J.-P. Rosay, "Fonctions plurisousharmoniques d'exhaustion bornées et domaines taut", *Math. Ann.* **257**:2 (1981), 171–184. MR 83g:32019 Zbl 0451.32012

[Kiernan 1970] P. Kiernan, "On the relations between taut, tight and hyperbolic manifolds", *Bull. Amer. Math. Soc.* **76** (1970), 49–51. MR 40 #5896 Zbl 0192.44103

[Koecher 1999] M. Koecher, *The Minnesota notes on Jordan algebras and their applications*, Lecture Notes in Mathematics **1710**, Springer, Berlin, 1999. MR 2001e:17040 Zbl 1072.17513

[Korányi 1969] A. Korányi, "Harmonic functions on Hermitian hyperbolic space", *Trans. Amer. Math. Soc.* **135** (1969), 507–516. MR 43 #3480 Zbl 0174.38801

[Łojasiewicz 1959] S. Łojasiewicz, "Sur le problème de la division", *Studia Math.* **18** (1959), 87–136. MR 21 #5893 Zbl 0115.10203

[Loos 1977] O. Loos, "Bounded symmetric domains and Jordan pairs", Mathematical Lectures, University of California, Irvine, 1977.

[Mok and Tsai 1992] N. Mok and I. H. Tsai, "Rigidity of convex realizations of irreducible bounded symmetric domains of rank ≥ 2", *J. Reine Angew. Math.* **431** (1992), 91–122. MR 93h:32046 Zbl 0765.32017

[Narasimhan 1971] R. Narasimhan, *Several complex variables*, The University of Chicago Press, Chicago, IL, 1971. MR 49 #7470 Zbl 0223.32001

[Roos 2000] G. Roos, "Jordan triple systems", pp. 425–536 in *Analysis and geometry on complex homogeneous domains*, Progress in Mathematics **185**, Birkhäuser, Boston, MA, 2000. MR 2001f:32036 Zbl 1043.17017

[Rudin 1980] W. Rudin, *Function theory in the unit ball of* $\mathbf{C}^n$, Grundlehren der Mathematischen Wissenschaften **241**, Springer, New York, 1980. MR 82i:32002 Zbl 0495.32001

[Tsai 1993] I. H. Tsai, "Rigidity of proper holomorphic maps between symmetric domains", *J. Differential Geom.* **37**:1 (1993), 123–160. MR 93m:32038 Zbl 0799.32027

[Tu 2002] Z.-H. Tu, "Rigidity of proper holomorphic mappings between equidimensional bounded symmetric domains", *Proc. Amer. Math. Soc.* **130**:4 (2002), 1035–1042. MR 2003a:32027 Zbl 0999.32007

[Tumanov and Khenkin 1982] A. E. Tumanov and G. M. Khenkin, "Local characterization of analytic automorphisms of classical domains", *Dokl. Akad. Nauk SSSR* **267**:4 (1982), 796–799. In Russian; translated in *Sov. Math. Dokl.* **26** (1982), 702–705. MR 85b:32048 Zbl 0529.32014

[Vigué 1991] J.-P. Vigué, "Un lemme de Schwarz pour les domaines bornés symétriques irréductibles et certains domaines bornés strictement convexes", *Indiana Univ. Math. J.* **40**:1 (1991), 293–304. MR 92c:32026 Zbl 0733.32025

[Wolf 1972] J. A. Wolf, "Fine structure of Hermitian symmetric spaces", pp. 271–357 in *Symmetric spaces: Short courses* (St. Louis, MO, 1969–1970), edited by W. M. Boothby and G. L. Weiss, Pure and App. Math. **8**, Dekker, New York, 1972. MR 53 #8516 Zbl 0257.32014

[Wolf and Korányi 1965] J. A. Wolf and A. Korányi, "Generalized Cayley transformations of bounded symmetric domains", *Amer. J. Math.* **87** (1965), 899–939. MR 33 #229 Zbl 0137.27403

Received June 10, 2013.

GAUTAM BHARALI
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE
BANGALORE 560012
INDIA
bharali@math.iisc.ernet.in

JAIKRISHNAN JANARDHANAN
DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF SCIENCE
BANGALORE 560012
INDIA
jaikrishnan@math.iisc.ernet.in

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.25

AN EXPLICIT MAJORANA REPRESENTATION OF THE GROUP 3^2:2 OF $3C$-PURE TYPE

HSIAN-YANG CHEN AND CHING HUNG LAM

We study a coset vertex operator algebra (VOA) $\widetilde{W}$ in the lattice VOA $V_{E_8^3}$. We show that the coset VOA $\widetilde{W}$ is generated by nine Ising vectors such that any two Ising vectors generate a $3C$ subVOA U_{3C}, and the group generated by the corresponding Miyamoto involutions has shape 3^2:2. This gives an explicit example for Majorana representations of the group 3^2:2 of $3C$-pure type.

1. Introduction

A vertex operator algebra (VOA) $V = \bigoplus_{n=0}^{\infty} V_n$ is said to be of *moonshine type* if $\dim(V_0) = 1$ and $V_1 = 0$. In this case, the weight-2 subspace V_2 has a commutative nonassociative product defined by $a \cdot b = a_1 b$ for $a, b \in V_2$ and it has a symmetric invariant bilinear form $\langle \cdot, \cdot \rangle$ given by $\langle a, b \rangle \mathbb{1} = a_3 b$ for $a, b \in V_2$ [Frenkel et al. 1988]. The algebra $(V_2, \cdot, \langle \cdot, \cdot \rangle)$ is often called the *Griess algebra* of V. An element $e \in V_2$ is called an *Ising vector* if $e \cdot e = 2e$ and the subVOA generated by e is isomorphic to the simple Virasoro VOA $L(\frac{1}{2}, 0)$ of central charge $\frac{1}{2}$. In [Miyamoto 1996], the basic properties of Ising vectors have been studied. Miyamoto also gave a simple method to construct involutive automorphisms of a VOA V from Ising vectors. These automorphisms are often called Miyamoto involutions. When V is the famous Moonshine VOA $V^\natural$, Miyamoto [2004] showed that there is a one-to-one correspondence between the $2A$-involutions of the Monster group and Ising vectors in $V^\natural$ (see also [Höhn 2010]). This correspondence is very useful for studying some mysterious phenomena of the Monster group and many problems about $2A$-involutions in the Monster group may also be translated into questions about Ising vectors. For example, McKay's observation on the affine E_8-diagram was studied in [Lam et al. 2007] using Miyamoto involutions and certain VOAs generated by two Ising vectors were constructed. Nine VOAs were constructed, denoted by U_{1A}, U_{2A}, U_{2B}, U_{3A}, U_{3C}, U_{4A}, U_{4B}, U_{5A}, and U_{6A} because of their connection to the 6-transposition property of the Monster group (see [ibid., Introduction]),

Partially supported by NSC grant 100-2628-M-001005-MY4 .
MSC2010: primary 17B69; secondary 20B25.
Keywords: vertex operator algebras, Ising vectors, Majorana representation.

where $1A, 2A, \ldots, 6A$ are the labels for certain conjugacy classes of the Monster as denoted in [Conway et al. 1985]. In [Sakuma 2007], Griess algebras generated by two Ising vectors contained in a moonshine-type VOA over $\mathbb{R}$ with a positive definite invariant form are classified. There are also nine possible cases, and they correspond exactly to the Griess algebras $\mathcal{G}U_{nX}$ of the nine VOAs U_{nX}, for nX in $\{1A, 2A, 2B, 3A, 3C, 4A, 4B, 5A, 6A\}$. Therefore, there is again a correspondence between the dihedral subgroups generated by two $2A$-involutions, up to conjugacy and the Griess subalgebras generated by two Ising vectors in $V^\natural$, up to isomorphism. It is also conjectured that the subVOA generated by two Ising vectors is isomorphic to one of the U_{nX}, for $nX \in \{1A, 2A, 2B, 3A, 3C, 4A, 4B, 5A, 6A\}$. However, this conjecture is still open except for the cases $1A$, $2A$, $2B$, $3A$, and $4B$.

Motivated by [Sakuma 2007], Ivanov [2009] axiomatized the properties of Ising vectors and introduced the notion of Majorana representations for finite groups. Ivanov and his research group also initiated a program on classifying the Majorana representations for various finite groups [Ivanov et al. 2010; Ivanov 2011a; 2011b; Ivanov and Seress 2012]. In particular, the famous 196884-dimensional Monster Griess algebra constructed by Griess [1982] is a Majorana representation of the Monster simple group. In fact, most known examples of Majorana representations are constructed as certain subalgebras of this Monster Griess algebra.

In this article, we construct explicitly a moonshine-type VOA $\widetilde{W}$ in the lattice VOA $V_{E_8^3}$. We show that the VOA $\widetilde{W}$ is generated by nine Ising vectors such that (1) any two of them generate a $3C$ subVOA U_{3C}; and (2) the group generated by the corresponding Miyamoto involutions has the shape $3^2{:}2$. Thus, we obtain an example for a Majorana representation of the group $3^2{:}2$ of $3C$-pure type. Recall that the centralizer of a $3C$-element in the Monster is isomorphic to $3 \times \mathrm{Th}$, where Th is the Thompson simple group [Conway et al. 1985]. The Thompson group Th has exactly three conjugacy classes of order 3 and by the character table, one can show that the Th conjugacy classes $3A$, $3B$, $3C$ are of the classes $3A$, $3B$, $3B$ in the Monster, respectively (see [ibid.] and [Wilson 1988, Section 4]). Therefore, there are no $3C$-pure 3^2 subgroups in the Monster and hence the VOA that we constructed cannot be embedded into the Moonshine VOA.

Our method is essentially a combination of the construction of the so-called dihedral subVOA from [Lam et al. 2007] and the construction of EE_8 pairs from [Griess and Lam 2011]. In fact, it is quite straightforward to find Ising vectors satisfying our hypotheses. The main difficulty is to show that the subVOA generated by these Ising vectors has zero weight-1 subspace.

The organization of this article is as follows. In Section 2, we recall some basic definitions and notation. We also review the structure of the so-called $3C$-algebra from [Lam et al. 2005; 2007]. In Section 3, we give an explicit construction of a coset subVOA $\widetilde{W}$ in the lattice $V_{E_8^3}$. We also construct explicitly several

Ising vectors satisfying our main hypotheses and show that the subVOA W they generate is of moonshine type. In Section 4, we show that the VOA W is isomorphic to the commutant subVOA $\widetilde{W} = \mathrm{Com}_{V_{E_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0))$ using the theory of parafermion VOA. The decomposition of W as a sum of irreducible modules of the parafermion VOA $K(\mathfrak{sl}_3(\mathbb{C}), 9)$ is also obtained. In Section 5, we give several structural results about Griess algebras generated by Ising vectors. We show that the Griess algebra generated by Ising vectors such that the subgroup generated by the corresponding Miyamoto involutions has the shape $3^2{:}2$ and is of $3C$-pure type is uniquely determined, up to isomorphisms. We also show that the VOA generated by these Ising vectors has central charge 4 and has a full subVOA isomorphic to $L(\frac{1}{2},0) \otimes L(\frac{21}{22},0) \otimes L(\frac{28}{11},0)$. In the Appendix, we explain several results which are used to show that $\dim(\widetilde{W}_2) = 9$.

2. Preliminaries

First we will recall some definitions and review several basic facts.

Definition 2.1. Let V be a VOA. A bilinear $\langle\langle\,\cdot\,,\,\cdot\,\rangle\rangle$ form on V is said to be *invariant* (or *contragredient*; see [Frenkel et al. 1993]) if

$$\langle\langle Y(a,z)u, v\rangle\rangle = \langle\langle u, Y(e^{zL(1)}(-z^{-2})^{L(0)}a, z^{-1})v\rangle\rangle \tag{2-1}$$

for any $a, u, v \in V$.

Definition 2.2. Let V be a VOA over $\mathbb{C}$. A *real form* of V is a subVOA $V_{\mathbb{R}}$ of V over $\mathbb{R}$ (with the same vacuum and Virasoro elements) such that $V = V_{\mathbb{R}} \otimes \mathbb{C}$. A real form $V_{\mathbb{R}}$ is said to be *positive definite* if the invariant form $\langle\langle\,\cdot\,,\,\cdot\,\rangle\rangle$ restricted to $V_{\mathbb{R}}$ is real-valued and positive definite.

Definition 2.3. Let V be a VOA. An element $v \in V_2$ is called a *simple Virasoro vector* of central charge c if the subVOA $\mathrm{Vir}(v)$ generated by e is isomorphic to the simple Virasoro VOA $L(c,0)$ of central charge c.

Definition 2.4. A simple Virasoro vector of central charge $\frac{1}{2}$ is called an *Ising vector*.

Remark 2.5. It is well known that the VOA $L(\frac{1}{2},0)$ is rational and has exactly three irreducible modules $L(\frac{1}{2},0)$, $L(\frac{1}{2},\frac{1}{2})$, and $L(\frac{1}{2},\frac{1}{16})$ (see [Dong et al. 1994; Miyamoto 1996]).

Remark 2.6. Let V be a VOA and let $e \in V$ be an Ising vector. Then we have the decomposition

$$V = V_e(0) \oplus V_e(\tfrac{1}{2}) \oplus V_e(\tfrac{1}{16}),$$

where $V_e(h)$ denotes the sum of all irreducible $\mathrm{Vir}(e)$-submodules of V isomorphic to $L(\frac{1}{2},h)$ for $h \in \{0, \frac{1}{2}, \frac{1}{16}\}$.

Theorem 2.7 [Miyamoto 1996]. *The linear map $\tau_e : V \to V$ defined by*

$$\tau_e := \begin{cases} 1 & \text{on } V_e(0) \oplus V_e(\frac{1}{2}), \\ -1 & \text{on } V_e(\frac{1}{16}), \end{cases} \tag{2-2}$$

is an automorphism of V.

Remark 2.8. On the fixed point subspace V^{τ_e} of τ_e, we have $V^{\tau_e} = V_e(0) \oplus V_e(\frac{1}{2})$. The linear map $\sigma_e : V^{\tau_e} \to V^{\tau_e}$ which acts as 1 on $V_e(0)$ and -1 on $V_e(\frac{1}{2})$ also defines an automorphism of V^{τ_e} [ibid.]. Nevertheless, we do not need this fact in this article.

The $3C$-algebra. We recall the properties of the $3C$-algebra U_{3C} from [Lam et al. 2005, Section 3.9] (see also [Sakuma 2007]).

Lemma 2.9. *Let $U = U_{3C}$ be the $3C$-algebra. Then*:

(1) $U_1 = 0$ *and U is generated by its weight-2 subspace U_2 as a VOA.*

(2) $\dim U_2 = 3$ *and it is spanned by three Ising vectors.*

(3) *There exist exactly three Ising vectors in U_2, say, e^0, e^1, e^2. Moreover, we have*

$$(e^i)_1(e^j) = \tfrac{1}{32}(e^i + e^j - e^k) \quad \text{and} \quad \langle e^i, e^j \rangle = \tfrac{1}{2^8}$$

for $i \neq j$ and $\{i, j, k\} = \{0, 1, 2\}$.

(4) *Let $g = \tau_{e^0}\tau_{e^1}$. Then g has order 3. Moreover, $e^1 = ge^0$ and $e^2 = g^2e^0 = ge^1$.*

(5) *The Virasoro element of U is given by*

$$\tfrac{32}{33}(e^0 + e^1 + e^2).$$

(6) *Let $a = \frac{32}{33}(e^0 + e^1 + e^2) - e^0$. Then a is a simple Virasoro vector of central charge $\frac{21}{22}$. Moreover, the subVOA generated by e^0 and a is isomorphic to $L(\frac{1}{2}, 0) \otimes L(\frac{21}{22}, 0)$.*

3. Commutant subVOAs in $V_{E_8 \perp E_8 \perp E_8}$

In this section, we shall construct explicitly a VOA $\widetilde{W}$ inside the lattice VOA $V_{E_8 \perp E_8 \perp E_8}$ such that (1) $\widetilde{W}$ is generated by nine Ising vectors and any two Ising vectors generate a $3C$ subVOA U_{3C}; and (2) the group generated by the corresponding Miyamoto involutions has the shape $3^2{:}2$.

Our notation for the lattice vertex operator algebra

$$V_L = M(1) \otimes \mathbb{C}\{L\} \tag{3-1}$$

associated with a positive definite even lattice L is standard [Frenkel et al. 1988]. In particular, $\mathfrak{h} = \mathbb{C} \otimes_{\mathbb{Z}} L$ is an abelian Lie algebra and we extend the bilinear form to $\mathfrak{h}$ by $\mathbb{C}$-linearity. Also, $\hat{\mathfrak{h}} = \mathfrak{h} \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}k$ is the corresponding affine algebra and $\mathbb{C}k$ is the one-dimensional center of $\hat{\mathfrak{h}}$. The subspace $M(1)$ given by $\mathbb{C}[\alpha_i(n) \mid 1 \le i \le d, n < 0]$ for a basis $\{\alpha_1, \dots, \alpha_d\}$ of $\mathfrak{h}$, where $\alpha(n) = \alpha \otimes t^n$, is the unique irreducible $\hat{\mathfrak{h}}$-module such that $\alpha(n) \cdot 1 = 0$ for all $\alpha \in \mathfrak{h}$ and n nonnegative, and $k = 1$. Also, $\mathbb{C}\{L\} = \mathrm{span}\{e^\beta \mid \beta \in L\}$ is the twisted group algebra of the additive group L such that $e^\beta e^\alpha = (-1)^{\langle \alpha, \beta \rangle} e^\alpha e^\beta$ for any $\alpha, \beta \in L$. The vacuum vector $\mathbb{1}$ of V_L is $1 \otimes e^0$ and the Virasoro element ω_L is $\frac{1}{2} \sum_{i=1}^d \beta_i(-1)^2 \cdot \mathbb{1}$ where $\{\beta_1, \dots, \beta_d\}$ is an orthonormal basis of $\mathfrak{h}$. For the explicit definition of the corresponding vertex operators, we shall refer to [ibid.] for details.

Definition 3.1. Let A and B be integral lattices with the inner products $\langle\ ,\ \rangle_A$ and $\langle\ ,\ \rangle_B$, respectively. The *tensor product of the lattices* A and B is defined to be the integral lattice which is isomorphic to $A \otimes_{\mathbb{Z}} B$ as a $\mathbb{Z}$-module and has the inner product given by

$$\langle \alpha \otimes \beta, \alpha' \otimes \beta' \rangle = \langle \alpha, \alpha' \rangle_A \cdot \langle \beta, \beta' \rangle_B, \quad \text{for any } \alpha, \alpha' \in A,\ \beta, \beta' \in B.$$

We simply denote the tensor product of the lattices A and B by $A \otimes B$.

$\sqrt{2}E_8$-sublattices. Let $L = E_8 \perp E_8 \perp E_8$ be the orthogonal sum of 3 copies of the root lattice of type E_8. Set

$$\begin{aligned} M &= \{(\alpha, -\alpha, 0) \mid \alpha \in E_8\} < L, \\ N &= \{(0, \alpha, -\alpha) \mid \alpha \in E_8\} < L. \end{aligned} \tag{3-2}$$

Then $M \cong N \cong \sqrt{2}E_8$ and $M + N \cong A_2 \otimes E_8$ (see [Griess and Lam 2011]). We also define

$$E := \mathrm{Ann}_L(M + N) = \{\beta \in L \mid \langle \beta, \beta' \rangle = 0 \text{ for all } \beta' \in M + N\}. \tag{3-3}$$

Note that $E = \{(\alpha, \alpha, \alpha) \mid \alpha \in E_8\} < L$ and there is a third $\sqrt{2}E_8$-sublattice

$$\widetilde{N} = \{(\alpha, 0, -\alpha) \mid \alpha \in E_8\} < M + N.$$

We shall fix a (bilinear) 2-cocycle $\varepsilon_0 : E_8 \times E_8 \to \mathbb{Z}_2$ such that

$$\begin{aligned} \varepsilon_0(\alpha, \alpha) &\equiv \tfrac{1}{2}\langle \alpha, \alpha \rangle \mod 2, \\ \varepsilon_0(\alpha, \beta) - \varepsilon_0(\beta, \alpha) &\equiv \langle \alpha, \beta \rangle \mod 2, \end{aligned} \tag{3-4}$$

for all $\alpha, \beta \in E_8$. Note that such a 2-cocycle exists (see [Frenkel et al. 1988, (6.1.27)–(6.1.29)]). Moreover, $e^\alpha e^{-\alpha} = -e^0$ for any $\alpha \in E_8$ such that $\langle \alpha, \alpha \rangle = 2$.

We shall extend ε_0 to L by defining

$$\varepsilon_0((\alpha, \alpha', \alpha''), (\beta, \beta', \beta'')) = \varepsilon_0(\alpha, \beta) + \varepsilon_0(\alpha', \beta') + \varepsilon_0(\alpha'', \beta'').$$

It is easy to check by direct calculations that ε_0 is trivial on M, N, or $\widetilde{N}$.

Affine vertex operator algebras. We recall the notion of affine vertex operator algebras [Frenkel and Zhu 1992; Dong and Lepowsky 1993]. Let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra and $\hat{\mathfrak{g}}$ the affine Kac–Moody Lie algebra associated with $\mathfrak{g}$. Let $\Pi = \{\alpha_1, \dots, \alpha_n\}$ be a set of simple roots and θ the highest root. Let Q be the root lattice of $\mathfrak{g}$. For any positive integer k, we set

$$P_+^k(\mathfrak{g}) = \{\Lambda \in \mathbb{Q} \otimes_{\mathbb{Z}} Q \mid \langle \alpha_i, \Lambda \rangle \in \mathbb{Z}_{\geq 0} \text{ for all } i = 1, \dots, n \text{ and } \langle \theta, \Lambda \rangle \leq k\},$$

the set of dominant integral weights for $\mathfrak{g}$ with level k.

Let $L_{\hat{\mathfrak{g}}}(k, \Lambda)$ be the irreducible module of $\hat{\mathfrak{g}}$ with highest weight Λ and level k. Then $L_{\hat{\mathfrak{g}}}(k, 0)$ forms a simple VOA with the Virasoro element given by the Sugawara construction

$$\Omega_{\mathfrak{g},k} = \frac{1}{2(k + h^\vee)} \sum (u_i)_{-1} u^i, \tag{3-5}$$

where $h^\vee$ is the dual Coxeter number, $\{u_i\}$ is a basis of $\mathfrak{g}$ and $\{u^i := (u_i)^*\}$ is the dual basis of $\{u_i\}$ with respect to the normalized Killing form (see [Frenkel and Zhu 1992]). Moreover, the central charge of $L_{\hat{\mathfrak{g}}}(k, 0)$ is

$$\frac{k \dim \mathfrak{g}}{k + h^\vee}. \tag{3-6}$$

A commutant subVOA. Consider the lattice VOA

$$V_L \cong V_{E_8} \otimes V_{E_8} \otimes V_{E_8}$$

and let $\boldsymbol{a}$ be an element of E_8 such that

$$K := \{\beta \in E_8 \mid \langle \beta, \boldsymbol{a} \rangle \in 3\mathbb{Z}\} \cong A_8.$$

Then, we have an embedding

$$V_{K \perp K \perp K} \cong V_K \otimes V_K \otimes V_K \hookrightarrow V_L.$$

It is also well known that $V_K \cong V_{A_8}$ is an irreducible level-1 representation of the affine Lie algebra $\widehat{\mathfrak{sl}}_9(\mathbb{C})$ [Frenkel et al. 1988]. Moreover, the weight-1 subspace $(V_K)_1$ is a simple Lie algebra isomorphic to $\mathfrak{sl}_9(\mathbb{C})$.

Let $\eta_i : K \to K \perp K \perp K$, $i = 1, 2, 3$, be the embedding of K into the i-th direct summand of $K \perp K \perp K$, i.e.,

$$\eta_1(\alpha) = (\alpha, 0, 0), \quad \eta_2(\alpha) = (0, \alpha, 0), \quad \eta_3(\alpha) = (0, 0, \alpha),$$

for any $\alpha \in K$.

Notation 3.2. For any $\alpha \in K(2) := \{\alpha \in K \mid \langle \alpha, \alpha \rangle = 2\}$, set

$$H_\alpha = (\alpha, \alpha, \alpha)(-1) \cdot \mathbb{1},$$
$$E_\alpha = e^{\eta_1(\alpha)} + e^{\eta_2(\alpha)} + e^{\eta_3(\alpha)}.$$

Then $\{H_\alpha, E_\alpha \mid \alpha \in K(2)\}$ generates a subVOA isomorphic to the affine VOA $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$ in V_L (see [Frenkel and Zhu 1992; Dong and Lepowsky 1993, Proposition 13.1]). Moreover, the Virasoro element of $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$ is given by

$$\Omega = \frac{1}{2(3+9)}\left[\sum_{k=1}^{8}(h^k, h^k, h^k)(-1)^2\cdot \mathbb{1} + \sum_{\alpha\in K(2)}(E_\alpha)_{-1}(-E_{-\alpha})\right],$$

where $\{h^1,\dots,h^8\}$ is an orthonormal basis of $K\otimes\mathbb{C} = E_8\otimes\mathbb{C}$. Note that the dual vector of E_α is $-E_{-\alpha}$.

Lemma 3.3. *Let M, N and E be defined as in* (3-2) *and* (3-3) *and denote the Virasoro element of a lattice VOA V_S by ω_S. Then we have*

$$\Omega = \omega_E + \tfrac{3}{4}\omega_{M+N} - \tfrac{1}{12}\sum_{\substack{\alpha\in K(2)\\ 1\le i,j\le 3,\, i\ne j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}.$$

Proof. Let $\{h^1,\dots,h^8\}$ be an orthonormal basis of $A_8\otimes\mathbb{C} = E_8\otimes\mathbb{C}$. Then

$$\begin{aligned}\Omega &= \frac{1}{2(3+9)}\Bigg[\sum_{k=1}^{8}(h^k, h^k, h^k)(-1)^2\cdot\mathbb{1}\\ &\qquad - \sum_{\alpha\in K(2)}\big(e^{\eta_1(\alpha)}+e^{\eta_2(\alpha)}+e^{\eta_3(\alpha)}\big)_{-1}\big(e^{-\eta_1(\alpha)}+e^{-\eta_2(\alpha)}+e^{-\eta_3(\alpha)}\big)\Bigg]\\ &= \frac{1}{24}\Bigg[6\omega_E + \sum_{\alpha\in K(2)}\sum_{i=1}^{3}\tfrac{1}{2}(\eta_i(\alpha)(-2)\cdot\mathbb{1}+\eta_i(\alpha)(-1)^2\cdot\mathbb{1}) - 2\sum_{\substack{\alpha\in K(2)\\ 1\le i,j\le 3\\ i\ne j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}\Bigg]\\ &= \tfrac{1}{4}\omega_E + \tfrac{18}{24}\omega_L - \tfrac{1}{12}\sum_{\substack{\alpha\in K(2)\\ 1\le i,j\le 3,\, i\ne j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}.\end{aligned}$$

Since $\omega_L = \omega_{M+N} + \omega_E$, we have

$$\Omega = \omega_E + \tfrac{3}{4}\omega_{M+N} - \tfrac{1}{12}\sum_{\substack{\alpha\in K(2)\\ 1\le i,j\le 3,\, i\ne j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}$$

as desired. □

Theorem 3.4. *Let*

$$\widetilde{W} = \mathrm{Com}_{V_L}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) = \{v\in V_L \mid x_n v = 0 \text{ for all } x\in L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0),\, n\ge 0\}$$

be the commutant subVOA of $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$ in V_L. Then the central charge of $\widetilde{W}$ is 4. *Moreover*, $\widetilde{W}_1 = 0$.

Proof. By (3-6), the central charge of $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$ is $3(80)/(3+9) = 20$. Hence, the central charge of $\widetilde{W} = \mathrm{Com}_{V_L}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0))$ is 4 $(= 24-20)$.

We now show that $\widetilde{W}_1 = 0$. Since $h(-1)\cdot\mathbb{1} \in L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$ for all $h \in E$,

$$\widetilde{W} = \mathrm{Com}_{V_L}\big(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)\big) \subset V_{M+N}.$$

Therefore, it suffices to show $\widetilde{W} \cap (V_{M+N})_1 = 0$.

Recall that $M + N \cong A_2 \otimes E_8$ has no roots. Thus,

$$(V_{M+N})_1 = \mathrm{span}_{\mathbb{C}}\{h(-1)\cdot\mathbb{1} \mid h \in (M+N)\otimes\mathbb{C}\}.$$

However, by Lemma 3.3,

$$\Omega_1 h(-1)\cdot\mathbb{1} = \Bigg(\omega_E + \tfrac{3}{4}\omega_{M+N} - \tfrac{1}{12}\sum_{\substack{\alpha\in K(2)\\ 1\le i,j\le 3,\, i\ne j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}\Bigg)_1 h(-1)\cdot\mathbb{1} = \tfrac{3}{4}h(-1)\cdot\mathbb{1} \neq 0$$

for any $0 \neq h \in (M+N)\otimes\mathbb{C}$. Thus, $\widetilde{W} \cap (V_{M+N})_1 = 0$ and we have $\widetilde{W}_1 = 0$. □

Ising vectors. Next we shall define explicitly some Ising vectors in V_L.

Definition 3.5. Let $\boldsymbol{a}$ be an element of E_8 such that

$$K = \{\beta \in E_8 \mid \langle \beta, \boldsymbol{a}\rangle \in 3\mathbb{Z}\} \cong A_8.$$

Set $\tilde{\boldsymbol{a}} = (\boldsymbol{a}, -\boldsymbol{a}, 0)$ and define an automorphism ρ of V_L by

$$\rho = \exp\Big(\frac{2\pi i}{3}\tilde{\boldsymbol{a}}(0)\Big).$$

Then ρ has order 3 and the fixed point subspace $V_M^{\rho} \cong V_{\sqrt{2}A_8}$.

Notation 3.6. Let M and N be defined as in (3-2). Set

$$\begin{aligned} e &:= e_M = \tfrac{1}{16}\omega_M + \tfrac{1}{32}\sum_{\alpha\in M(4)} e^{\alpha},\\ f &:= e_N = \tfrac{1}{16}\omega_N + \tfrac{1}{32}\sum_{\alpha\in N(4)} e^{\alpha},\\ e_{\widetilde{N}} &:= \tfrac{1}{16}\omega_{\widetilde{N}} + \tfrac{1}{32}\sum_{\alpha\in \widetilde{N}(4)} e^{\alpha},\\ e' &:= \rho(e). \end{aligned}$$

It is shown in [Dong et al. 1998] that e, f and $e_{\widetilde{N}}$ are Ising vectors and hence $e' = \rho(e)$ is also an Ising vector (see also [Lam et al. 2005; 2007]).

The following lemma can be proved by direct calculations (see [Lam et al. 2005; 2007; Griess and Lam 2011]).

Lemma 3.7. *We have* $\langle e, f\rangle = \langle e, e'\rangle = \langle f, e'\rangle = 1/2^8$. *Moreover, the subVOAs* $\mathrm{VOA}(e, f)$, $\mathrm{VOA}(e, g)$, $\mathrm{VOA}(f, g)$ *generated by* $\{e, f\}$, $\{e, e'\}$, *and* $\{f, e'\}$, *are isomorphic to the* $3C$*-algebra* U_{3C}. *We also have* $e_M \cdot e_N = \frac{1}{32}(e_M + e_N - e_{\widetilde{N}})$, *and hence* $\tau_e(f) = e_{\widetilde{N}}$.

Notation 3.8. Let $W := \mathrm{VOA}(e, f, e')$ be the subVOA generated by e, f, and e'. We also denote $h = \tau_e\tau_f$ and $g = \tau_e\tau_{e'}$. Then g and h both have order 3. Note also that $e, f, e' \in V_{M+N}$ and thus $W < V_{M+N} \cong V_{A_2\otimes E_8}$.

Lemma 3.9. *The elements* g *and* h *commute as automorphisms of* W.

Proof. Recall that $g = \tau_e\tau_{e'} = \rho$ on V_L (see [Lam et al. 2007]). Also, $h(e) = f = e_N$ and $h^2(e) = e_{\widetilde{N}}$.

By a direct calculation, we have

$$hg(e) = hgh^{-1}h(e) = \rho^h(e_N),$$

where $\rho^h = h\rho h^{-1} = \exp\bigl(\frac{2\pi i}{3}(0, \boldsymbol{a}, -\boldsymbol{a})(0)\bigr)$.

Since $\langle(0, \beta, -\beta), (0, \boldsymbol{a}, -\boldsymbol{a})\rangle = 2\langle\beta, \boldsymbol{a}\rangle$ and $\langle(0, \beta, -\beta), (\boldsymbol{a}, -\boldsymbol{a}, 0)\rangle = -\langle\beta, \boldsymbol{a}\rangle$, we have

$$gh(e) = \rho(e_N) = \rho^h(e_N) = hg(e).$$

Similarly, we have

$$hg(e') = hg^2(e) = (\rho^h)^2(e_N), \quad gh(e') = ghg(e) = g(\rho^h(e_N)) = (\rho^h)^2(e_N)$$

and

$$hg(f) = hgh(e) = (hgh^2)h^2(e) = \rho^h(e_{\widetilde{N}}), \quad gh(f) = g(e_{\widetilde{N}}) = \rho(e_{\widetilde{N}}).$$

Hence $gh = hg$ on W. □

Notation 3.10. For any $0 \le i, j \le 2$, denote

$$e^{i,j} = g^i h^j(e).$$

In particular, we have

$$\begin{aligned} &e^{0,0} = e_M, && e^{0,1} = e_N, && e^{0,2} = e_{\widetilde{N}},\\ &e^{1,0} = \rho e_M, && e^{1,1} = \rho e_N, && e^{1,2} = \rho e_{\widetilde{N}},\\ &e^{2,0} = \rho^2 e_M, && e^{2,1} = \rho^2 e_N, && e^{2,2} = \rho^2 e_{\widetilde{N}}. \end{aligned}$$

Remark 3.11. By the same methods as in [Lam et al. 2007; Griess and Lam 2011], it is quite straightforward to verify that $\langle e^{i,j}, e^{i'j'}\rangle = \frac{1}{2^8}$ whenever $(i, j) \neq (i', j')$.

Lemma 3.12. *Let G be the subgroup of* $\mathrm{Aut}(W)$ *generated by τ_e, τ_f and $\tau_{e'}$. Then $G = \langle g, h\rangle : \langle \tau_e\rangle$, where $\langle g, h\rangle$ is elementary abelian of order* 3^2 *and τ_e inverts g and h.*

Proof. By Lemma 3.9, we know that the group $\langle g, h\rangle$ generated by g and h is elementary abelian of order 3^2. Also, τ_e inverts g and h because $\tau_e g \tau_e = \tau_e(\tau_e\tau_{e'})\tau_e = \tau_{e'}\tau_e = g^{-1}$ and $\tau_e h \tau_e = \tau_e(\tau_e\tau_f)\tau_e = \tau_f\tau_e = h^{-1}$.

First, we shall prove that $\langle g, h\rangle$ is normal in G. By Lemma 3.9 we have $gh = hg$ and hence $\tau_f\tau_e\tau_{e'} = \tau_{e'}\tau_e\tau_f$. Thus $\tau_f h\tau_f = \tau_f\tau_e\tau_{e'}\tau_f = \tau_{e'}\tau_e\tau_f^2 = \tau_{e'}\tau_e = h^2 \in \langle g, h\rangle$. Similar computation gives that $\langle g, h\rangle$ is normal in G.

Next we show that $G = \langle g, h\rangle\langle\tau_e\rangle$. Recall that τ_e, τ_f and $\tau_{e'}$ are involutions. Thus every nonidentity element in G has the form

$$\tau_{a_1}\tau_{a_2}\cdots\tau_{a_k},$$

where $a_i = e,\ f$, or e' and $a_i \neq a_{i+1}$ for $i = 1, \ldots, k-1$.

Note also that $\tau_e\tau_{e'} = g$, $\tau_e\tau_f = h$, $\tau_f\tau_{e'} = h^{-1}g$, and g and h have order 3. Hence, $\tau_a\tau_{a'} \in \langle g, h\rangle$ for any $a, a' \in \{e, f, e'\}$. Therefore, $\tau_{a_1}\tau_{a_2}\cdots\tau_{a_k} \in \langle g, h\rangle$ if k is even and $\tau_{a_1}\tau_{a_2}\cdots\tau_{a_k} = (\tau_{a_1}\tau_{a_2}\cdots\tau_{a_k}\tau_e)\tau_e \in \langle g, h\rangle\langle\tau_e\rangle$ if k is odd. Thus we have $G = \langle g, h\rangle\langle\tau_e\rangle$.

Since $|\langle g, h\rangle| = 3^2$ and $|\langle\tau_e\rangle| = 2$, we get $\langle g, h\rangle\cap\langle\tau_e\rangle = 1$. Hence $G = \langle g, h\rangle : \langle\tau_e\rangle$ as desired. □

Lemma 3.13. *Let Ω be the Virasoro element of $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3, 0)$. Then*

$$\Omega = \omega_L - \tfrac{8}{9}\sum_{0\le i,j\le 2} e^{i,j}.$$

Proof. By Lemma 3.3, we have

$$\Omega = \omega_E + \tfrac{3}{4}\omega_{M+N} - \tfrac{1}{12}\sum_{\substack{\alpha\in K(2)\\ 1\le i,j\le 3,\, i\neq j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}.$$

Now let us set $\Delta^i := \{\beta \in E_8(2) \mid \langle \boldsymbol{a}, \beta\rangle = i \mod 3\mathbb{Z}\}$ for $i = 0, 1, 2$. Note that $\Delta^0 = K(2)$. Then we have

$$e^{0,0} = e_M = \tfrac{1}{16}\omega_M + \tfrac{1}{32}\sum_{i=0}^{2}\sum_{\alpha\in\Delta^i} e^{(\alpha,-\alpha,0)},$$

$$e^{1,0} = \rho e_M = \tfrac{1}{16}\omega_M + \tfrac{1}{32}\sum_{i=0}^{2}\sum_{\alpha\in\Delta^i} \xi^{2i} e^{(\alpha,-\alpha,0)},$$

$$e^{2,0} = \rho^2 e_M = \tfrac{1}{16}\omega_M + \tfrac{1}{32}\sum_{i=0}^{2}\sum_{\alpha\in\Delta^i} \xi^{i} e^{(\alpha,-\alpha,0)}.$$

Hence

$$\sum_{i=0}^{2} e^{i,0} = \left(1+\rho+\rho^2\right)e_M = \tfrac{3}{16}\omega_M + \tfrac{3}{32}\sum_{\alpha\in K(2)} e^{(\alpha,-\alpha,0)}.$$

A similar computation gives

$$\sum_{0\le i,j\le 2} e^{i,j} = \tfrac{3}{16}(\omega_M+\omega_N+\omega_{\widetilde{N}}) + \tfrac{3}{32}\sum_{\substack{\alpha\in K(2)\\ 1\le i,j\le 3,\, i\ne j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}.$$

Recall that $M+N\cong A_2\otimes E_8$. It contains a full rank sublattice isometric to $(\sqrt{2}A_2)^8$ and hence ω_{M+N} is the sum of the conformal elements of each tensor copy of $V_{\sqrt{2}A_2}^{\otimes 8}$. We also note that the conformal element of the lattice VOA $V_{\sqrt{2}A_2}$ is given by

$$\begin{aligned}\omega_{\sqrt{2}A_2} &= \tfrac{1}{6}\left(\alpha_1(-1)^2+\alpha_2(-1)^2+\alpha_3(-1)^2\right)\cdot\mathbb{1}\\ &= \tfrac{2}{3}\left(\tfrac{1}{2}\left(\frac{\alpha_1(-1)}{\sqrt{2}}\right)^2+\tfrac{1}{2}\left(\frac{\alpha_2(-1)}{\sqrt{2}}\right)^2+\tfrac{1}{2}\left(\frac{\alpha_3(-1)}{\sqrt{2}}\right)^2\right)\cdot\mathbb{1},\end{aligned}$$

where $\alpha_1,\alpha_2,\alpha_3$ are positive roots of a root lattice type A_2 [Dong et al. 1998].

Thus $\omega_{M+N}=\frac{2}{3}(\omega_M+\omega_N+\omega_{\widetilde{N}})$ and we get

$$\Omega = \omega_L - \tfrac{8}{9}\sum_{0\le i,j\le 2} e^{i,j},$$

as desired. Note that $\omega_L=\omega_E+\omega_{M+N}$. ◻

Lemma 3.14. *For any* $0\le i,\, j\le 2$, *we have* $e^{i,j}\in\widetilde{W}=\mathrm{Com}_{V_L}\left(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)\right)$. *Hence* $W\subset\widetilde{W}$ *and* $W_1=0$.

Proof. Since $e^{i,j}\in V_{M+N}$ and $E=\{(\alpha,\alpha,\alpha)\mid \alpha\in E_8\}$ is orthogonal to $M+N$, it is clear that $(H_\alpha)_n e^{i,j}=0$ for all $n\ge 0$. It is also clear that $(E_\alpha)_n e^{i,j}=0$ for any root $\alpha\in K$ and $n\ge 2$.

Recall from [Frenkel et al. 1988] that

$$Y(e^\alpha,z)=\exp\left(\sum_{n\in\mathbb{Z}^+}\frac{\alpha(-n)}{n}z^n\right)\exp\left(\sum_{n\in\mathbb{Z}^+}\frac{\alpha(n)}{-n}z^{-n}\right)e^\alpha z^\alpha.$$

Now let $\sigma=(123)$ be a 3-cycle. Then by direct calculation, we have

$$\begin{aligned}(E_\alpha)_1 e^{i,j} &= (E_\alpha)_1(\rho^i h^j e_M)\\ &= \left(e^{\eta_1(\alpha)}+e^{\eta_2(\alpha)}+e^{\eta_3(\alpha)}\right)_1\\ &\qquad\times\left(\tfrac{1}{16}\omega_{h^j(M)}+\tfrac{1}{32}\sum_{\alpha\in\Delta^+(E_8)}\rho^i\left(e^{(\eta_{\sigma^j(1)}-\eta_{\sigma^j(2)})(\alpha)}+e^{-(\eta_{\sigma^j(1)}-\eta_{\sigma^j(2)})(\alpha)}\right)\right)\\ &= \tfrac{1}{16}\langle\alpha,\alpha\rangle^2\tfrac{1}{8}\left(e^{\eta_{\sigma^j(1)}(\alpha)}+e^{\eta_{\sigma^j(2)}(\alpha)}\right)+\tfrac{1}{32}\varepsilon(\alpha,-\alpha)\left(e^{\eta_{\sigma^j(1)}(\alpha)}+e^{\eta_{\sigma^j(2)}(\alpha)}\right)\\ &= 0,\end{aligned}$$

and

$$\begin{aligned}(E_\alpha)_0 e^{i,j} &= \big(e^{\eta_1(\alpha)} + e^{\eta_2(\alpha)} + e^{\eta_3(\alpha)}\big)_0 \\ &\qquad \times \left(\tfrac{1}{16}\omega_{h^j(M)} + \tfrac{1}{32}\sum_{\alpha\in\Delta^+(E_8)} \rho^i\big(e^{(\eta_{\sigma^j(1)}-\eta_{\sigma^j(2)})(\alpha)} + e^{-(\eta_{\sigma^j(1)}-\eta_{\sigma^j(2)})(\alpha)}\big)\right) \\ &= \tfrac{1}{16}\Big(\langle\alpha,\alpha\rangle^2 \tfrac{1}{8}\big(\eta_{\sigma^j(1)}(\alpha)(-1)e^{\eta_{\sigma^j(1)}(\alpha)} + \eta_{\sigma^j(2)}(\alpha)(-1)e^{\eta_{\sigma^j(2)}(\alpha)}\big) \\ &\qquad\qquad - 2\langle\alpha,\alpha\rangle \tfrac{1}{8}\big((\eta_{\sigma^j(1)} - \eta_{\sigma^j(2)})(\alpha)(-1)e^{\eta_{\sigma^j(1)}(\alpha)} \\ &\qquad\qquad\qquad - (\eta_{\sigma^j(1)} - \eta_{\sigma^j(2)})(\alpha)(-1)e^{\eta_{\sigma^j(2)}(\alpha)}\big)\Big) \\ &\qquad + \tfrac{1}{32}\varepsilon(\alpha,-\alpha)\big(\eta_{\sigma^j(2)}(\alpha)(-1)e^{\eta_{\sigma^j(1)}(\alpha)} + \eta_{\sigma^j(1)}(\alpha)(-1)e^{\eta_{\sigma^j(2)}(\alpha)}\big) \\ &= 0\end{aligned}$$

for any root $\alpha \in K$. Therefore, $(E_\alpha)_n e^{i,j} = 0$ for all $n \geq 0$. Since $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$ is generated by E_α and H_α, we have the desired conclusion. □

Remark 3.15. Note that the lattice VOA V_L also contains a subVOA isomorphic to $L_{\hat{E}_8}(3,0)$, the level-3 affine VOA associated to the Kac–Moody Lie algebra of type $E_8^{(1)}$. The central charge of $\mathrm{Com}_{V_L}(L_{\hat{E}_8}(3,0))$ is $\frac{16}{11}$, which is the same as U_{3C}. In fact, it can be shown by the similar calculation as Lemma 3.14 that e_M and e_N defined in Notation 3.6 are contained in $\mathrm{Com}_{V_L}(L_{\hat{E}_8}(3,0))$. Moreover,

$$U_{3C} \cong \mathrm{VOA}(e_M, e_N) = \mathrm{Com}_{V_L}(L_{\hat{E}_8}(3,0)).$$

4. Parafermion VOA and W

In this section, we shall show that the VOA W defined in Notation 3.8 is, in fact, equal to the commutant subVOA $\widetilde{W} = \mathrm{Com}_{V_L}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0))$. Recall that the lattice VOA $V_{A_3^8}$ contains a full subVOA $K(\mathfrak{sl}_3(\mathbb{C}),9) \otimes L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$ (see [Lam 2014]), where $K(\mathfrak{sl}_3(\mathbb{C}),9)$ is the parafermion VOA associated to the affine VOA $L_{\widehat{\mathfrak{sl}}_3(\mathbb{C})}(9,0)$. Therefore, the VOA $\widetilde{W}$ contains a full subVOA isomorphic to the parafermion VOA $K(\mathfrak{sl}_3(\mathbb{C}),9)$.

Parafermion VOA. First, we recall the definition of parafermion VOA from [Dong and Wang 2010], henceforth abbreviated [DW] (cf. [Dong et al. 2009; 2010]).

Let $\mathfrak{g}$ be a finite-dimensional simple Lie algebra and $\hat{\mathfrak{g}}$ the affine Kac–Moody Lie algebra associated with $\mathfrak{g}$. The level-k affine vertex operator algebra $L_{\hat{\mathfrak{g}}}(k,0)$ contains a Heisenberg vertex operator algebra corresponding to a Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$. Let $M_{\hat{\mathfrak{h}}}(k,0)$ be the vertex operator subalgebra of $L_{\hat{\mathfrak{g}}}(k,0)$ generated by $h(-1)\cdot\mathbb{1}$ for $h \in \mathfrak{h}$. The commutant $K(\mathfrak{g},k)$ of $M_{\hat{\mathfrak{h}}}(k,0)$ in $L_{\hat{\mathfrak{g}}}(k,0)$ is called a parafermion vertex operator algebra.

The VOA $L_{\hat{\mathfrak{g}}}(k,0)$ is completely reducible as an $M_{\hat{\mathfrak{h}}}(k,0)$-module and we have a decomposition (see [DW]).

Lemma 4.1. *For any $\lambda \in \mathfrak{h}^*$, let $M_{\hat{\mathfrak{h}}}(k,\lambda)$ be the irreducible highest weight module for $\hat{\mathfrak{h}}$ with a highest weight vector v_λ such that $h(0)v_\lambda = \lambda(h)v_\lambda$ for $h \in \mathfrak{h}$. Set*

$$K_{\mathfrak{g},k}(\lambda) = K_{\mathfrak{g},k}(0,\lambda) = \big\{ v \in L_{\hat{\mathfrak{g}}}(k,0) \;\big|\; h(m)v = \lambda(h)\delta_{m,0}v \text{ for } h \in \mathfrak{h}, m \geq 0 \big\}.$$

Then we have

$$L_{\hat{\mathfrak{g}}}(k,0) = \bigoplus_{\lambda \in Q} K_{\mathfrak{g},k}(\lambda) \otimes M_{\hat{\mathfrak{h}}}(k,\lambda),$$

where Q is the root lattice of $\mathfrak{g}$.

Similarly, for any dominant integral weight $\Lambda \in P_+^k(\mathfrak{g})$, we also have the decomposition.

Lemma 4.2. *Set*

$$K_{\mathfrak{g},k}(\Lambda,\lambda) = \big\{ v \in L_{\hat{\mathfrak{g}}}(k,\Lambda) \;\big|\; h(m)v = \lambda(h)\delta_{m,0}v \text{ for } h \in \mathfrak{h}, m \geq 0 \big\}.$$

Then

$$L_{\hat{\mathfrak{g}}}(k,\Lambda) = \bigoplus_{\lambda \in \Lambda + Q} K_{\mathfrak{g},k}(\Lambda,\lambda) \otimes M_{\hat{\mathfrak{h}}}(k,\lambda).$$

A generating set. In [DW], it is shown that the parafermion VOA $K(\mathfrak{g},k)$ is generated by subVOAs isomorphic to $K(\mathfrak{sl}_2(\mathbb{C}),k)$. We first give a brief review of their work.

Let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$ and let Δ_+ be the set of all positive roots of $\mathfrak{g}$. Then

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Delta_+} (\mathbb{C}x_\alpha \oplus \mathbb{C}x_{-\alpha}),$$

where $x_{\pm\alpha} \in \mathfrak{g}_{\pm\alpha} = \{u \in \mathfrak{g} \mid [h,u] = \pm\alpha(h)u \text{ for all } h \in \mathfrak{h}\}$.

Notation 4.3. For any $\alpha \in \Delta_+$, let $h_\alpha = [x_\alpha, x_{-\alpha}]$. Then $S_\alpha = \operatorname{span}\{h_\alpha, x_\alpha, x_{-\alpha}\}$ is a Lie subalgebra of $\mathfrak{g}$ isomorphic to $\mathfrak{sl}_2(\mathbb{C})$. Define

$$\omega_\alpha = \frac{1}{2k(k+2)}(kh_\alpha(-2)1 - h_\alpha(-1)2\mathbb{1} + 2kx_\alpha(-1)x_{-\alpha}(-1)\mathbb{1})$$

and

$$\begin{aligned} W_\alpha^3 &= k^2h_\alpha(-3)\mathbb{1} + 3kh_\alpha(-2)h_\alpha(-1)\mathbb{1} + 2h_\alpha(-1)^3\mathbb{1} \\ &\quad - 6kh_\alpha(-1)x_\alpha(-1)x_\alpha(-1)\mathbb{1} + 3k^2x_\alpha(-2)x_\alpha(-1)\mathbb{1} - 3k^2x_\alpha(-1)x_\alpha(-2)\mathbb{1}. \end{aligned}$$

We use P_α to denote the vertex operator subalgebra of $K(\mathfrak{g},k)$ generated by ω_α and W_α^3 for $\alpha \in \Delta_+$.

Theorem 4.4 [DW, Theorem 4.2]. *The simple vertex operator algebra $K(\mathfrak{g}, k)$ is generated by P_α, $\alpha \in \Delta_+$ and P_α is a simple vertex operator algebra isomorphic to the parafermion vertex operator algebra $K(\mathfrak{sl}_2(\mathbb{C}), k)$ associated to $\mathfrak{sl}_2(\mathbb{C})$.*

The lattice VOA $V_{A_n^{k+1}}$. Next we recall an embedding of the VOA

$$K(\mathfrak{sl}_{k+1}, n+1) \otimes L_{\widehat{\mathfrak{sl}}_{n+1}(\mathbb{C})}(k+1, 0)$$

into the lattice VOA $V_{A_n^{k+1}}$ from [Lam 2014].

We use the standard model for the root lattice of type A_ℓ. In particular,

$$A_\ell = \Big\{ \sum a_i \epsilon_i \in \mathbb{Z}^{\ell+1} \;\Big|\; a_i \in \mathbb{Z} \text{ and } \sum_{i=1}^{\ell+1} a_i = 0 \Big\},$$

where ϵ_i is the row vector whose i-th entry is 1 and all other entries are 0. The dual lattice

$$A_\ell^* = \bigcup_{i=0}^{\ell} (\gamma_{A_\ell}(i) + A_\ell),$$

where $\gamma_{A_\ell}(i) = \dfrac{1}{\ell+1}\Big(\sum_{j=1}^{\ell+1-i} i\epsilon_j - \sum_{j=\ell+1-i+1}^{\ell+1} (\ell+1-i)\epsilon_j \Big)$ for $i = 0, \ldots, \ell$.

Notation 4.5. Let n and k be positive integers. We shall consider two injective maps $\eta_i : \mathbb{Z}^{n+1} \to \mathbb{Z}^{(n+1)(k+1)}$ and $\iota_i : \mathbb{Z}^{k+1} \to \mathbb{Z}^{(n+1)(k+1)}$ defined by

$$\eta_i(\epsilon_j) = \epsilon_{(n+1)(i-1)+j} \quad \text{and} \quad \iota_i(\epsilon_j) = \epsilon_{(n+1)(j-1)+i}$$

for $i = 1, \ldots, k+1$, $j = 1, \ldots, n+1$.

Let

$$d_{k+1} = \sum_{j=1}^{k+1} \eta_j : \mathbb{Z}^{n+1} \to \mathbb{Z}^{(n+1)(k+1)} \quad \text{and} \quad \mu_{n+1} = \sum_{j=1}^{n+1} \iota_j : \mathbb{Z}^{k+1} \to \mathbb{Z}^{(n+1)(k+1)}.$$

Then we have

$$d_{k+1}(a_1, \ldots, a_{n+1}) = (a_1, \ldots, a_{n+1}, a_1, \ldots, a_{n+1}, \ldots, a_1, \ldots, a_{n+1}),$$
$$\mu_{n+1}(a_1, \ldots, a_{k+1}) = (a_1, \ldots, a_1, a_2, \ldots, a_2, \ldots, a_{k+1}, \ldots, a_{k+1}).$$

Set $X = d_{k+1}(A_n)$ and $Y = \mu_{n+1}(A_k)$. Then $X \cong \sqrt{k+1}A_n$ and $Y \cong \sqrt{n+1}A_k$. Moreover, we have

$$\begin{aligned} \operatorname{Ann}_{A_{(n+1)(k+1)-1}}(Y) &= \bigoplus_{i=1}^{k+1} \eta_i(A_n) \cong A_n^{k+1}, \\ \operatorname{Ann}_{A_{(n+1)(k+1)-1}}(X) &= \bigoplus_{j=1}^{n+1} \iota_j(A_k) \cong A_k^{n+1}, \end{aligned} \tag{4-1}$$

where $\mathrm{Ann}_A(B)=\{x\in A \mid \langle x, y\rangle=0 \text{ for all } y\in B\}$ is the annihilator of a sublattice B in an integral lattice A.

By the same construction as in Notation 3.2 (see also [Dong and Lepowsky 1993, Chapter 13]), one can obtain subVOAs isomorphic to $L_{\widehat{\mathfrak{sl}}_{n+1}(\mathbb{C})}(k+1,0)$ and $L_{\widehat{\mathfrak{sl}}_{k+1}(\mathbb{C})}(n+1,0)$ in the lattice VOA $V_{A_{(k+1)(n+1)-1}}$.

The next proposition is well known in the literature [Kac and Wakimoto 1988; Nakanishi and Tsuchiya 1992; Lam 2014].

Proposition 4.6. *The VOAs $L_{\widehat{\mathfrak{sl}}_{n+1}(\mathbb{C})}(k+1,0)$ and $L_{\widehat{\mathfrak{sl}}_{k+1}(\mathbb{C})}(n+1,0)$ are mutually commutative in the lattice VOA $V_{A_{(k+1)(n+1)-1}}$. Moreover,*

$$L_{\widehat{\mathfrak{sl}}_{n+1}(\mathbb{C})}(k+1,0)\otimes L_{\widehat{\mathfrak{sl}}_{k+1}(\mathbb{C})}(n+1,0)$$

is a full subVOA of $V_{(n+1)(k+1)-1}$.

Remark 4.7. It is also known that the VOA $V_{\mu_{n+1}(A_k)}$ is contained in the affine VOA $L_{\widehat{\mathfrak{sl}}_{k+1}(\mathbb{C})}(n+1,0)$ and $K(\mathfrak{sl}_{k+1}(\mathbb{C}),n+1)=\mathrm{Com}_{L_{\widehat{\mathfrak{sl}}_{k+1}(\mathbb{C})}(n+1,0)}(V_{\mu_{n+1}(A_k)})$ (see [Lam 2014, Lemma 4.1]). Moreover, for any $\Lambda\in P^+_{n+1}(\mathfrak{sl}_{k+1}(\mathbb{C}))$, we have the decomposition

$$L_{\widehat{\mathfrak{sl}}_{k+1}}(n+1,\Lambda)=\bigoplus_{\lambda\in\frac{\frac{1}{n+1}\mu_{n+1}(\Lambda+A_k)}{\mu_{n+1}(A_k)}} K_{\mathfrak{sl}_{k+1}(\mathbb{C}),n+1}(\Lambda,(n+1)\bar{\lambda})\otimes V_{\lambda+\mu_{n+1}(A_k)} \tag{4-2}$$

as a module of $V_{\mu_{n+1}(A_k)}\otimes K(\mathfrak{sl}_{k+1}(\mathbb{C}),n+1)$ ($\bar{\lambda}\in\frac{1}{n+1}A_k^*$) such that $\mu_{n+1}(\bar{\lambda})=\lambda$ (see [ibid., Lemma 4.3]).

Note that it is shown in [Dong and Lepowsky 1993, Theorem 14.20] that $K_{\mathfrak{sl}_{k+1}(\mathbb{C}),n+1}(\Lambda,(n+1)\bar{\lambda})$, for $\Lambda\in P_+^{n+1}(\mathfrak{sl}_{k+1}(\mathbb{C}))$, $\lambda\in(\mu_{n+1}(A_k))^*$, are irreducible $K(\mathfrak{sl}_{k+1}(\mathbb{C}),n+1)$-modules.

Next we consider the case $n=8$, $k=2$. Then $(n+1)(k+1)-1=26$. We shall study the decomposition of $\widetilde{W}=\mathrm{Com}_{V_{E_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0))$ as a $K(\mathfrak{sl}_3(\mathbb{C}),9)$-module.

Set

$$\nu_1=\eta_1-\eta_2,\qquad \nu_2=\eta_2-\eta_3,$$

and define $\mu=\mu_3:\mathbb{Z}^3\to\mathbb{Z}^{27}$ by

$$\mu(a_1,a_2,a_3)=(a_1,\dots,a_1,a_2,\dots,a_2,a_3,\dots a_3).$$

Note that $Y=\mu(A_2)\cong 3A_2$ and

$$\mathrm{Ann}_{A_{26}}(Y)=\{\alpha\in A_{26}\mid\langle\alpha,\beta\rangle=0 \text{ for any } \beta\in Y\}\cong A_8^3.$$

Next we discuss the coset decomposition $Y+A_8^3$ in A_{26}.

Lemma 4.8. *Let $\alpha_1 = (1, -1, 0)$ and $\alpha_2 = (0, 1, -1)$ be roots of A_2. Then we have*

$$A_{26} = \bigcup_{0\le i,j\le 8} \Big(\big(-\tfrac{1}{9}(i\mu(\alpha_1) + j\mu(\alpha_2)) + Y\big) + \big(\nu_1(\gamma_{A_8}(i)) + \nu_2(\gamma_{A_8}(j)) + A_8^3\big)\Big).$$

Proof. First we note that $[A_{26} : Y + A_8^3] = \sqrt{(9^2 \cdot 3) \cdot 9^3/27} = 9^2$. Moreover, we have

$$-\tfrac{1}{9}(i\mu(\alpha_1) + j\mu(\alpha_2)) + \nu_1(\gamma_{A_8}(i)) + \nu_2(\gamma_{A_8}(j)) = \sum_{k=10-i}^{9} \iota_k(\alpha_1) + \sum_{k'=10-i}^{9} \iota_{k'}(\alpha_2).$$

Note that $\sum_{k=10-i}^{9} \iota_k(\alpha_p) \notin Y + A_8^3$ for any $i \neq 0$, $p = 1, 2$. Therefore,

$$\big(-\tfrac{1}{9}(i\mu(\alpha_1) + j\mu(\alpha_2)) + Y\big) + \big(\nu_1(\gamma_{A_8}(i)) + \nu_2(\gamma_{A_8}(j)) + A_8^3\big),$$

for $i, j = 0, \dots, 8$, give 9^2 distinct cosets in $A_{26}/(Y + A_8^3)$. Thus, we have the desired conclusion. □

Lemma 4.9. *Let $\delta = \gamma_{A_8}(3) = \frac{1}{3}(1^6, (-2)^3) \in A_8^*$. Then for any $k, \ell = 0, \pm 1$, we have*

$$\begin{aligned} \mathrm{Com}_{V_{(k\nu_1+\ell\nu_2)(\delta)+A_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) &= \big\{v \in V_{(k\nu_1+\ell\nu_2)(\delta)+A_8^3} \bigm| \Omega_n v = 0 \textit{ for all } n \ge 0\big\} \\ &\cong K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, -3(k\alpha_1 + \ell\alpha_2)). \end{aligned}$$

Proof. By Lemma 4.8,

$$V_{A_{26}} = \bigoplus_{0\le i,j\le 8} V_{-\frac{1}{9}(i\mu(\alpha_1)+j\mu(\alpha_2))+Y} \otimes V_{\nu_1(\gamma_{A_8}(i))+\nu_2(\gamma_{A_8}(j))+A_8^3}.$$

Moreover, by (4-2),

$$L_{\widehat{\mathfrak{sl}}_3}(9,0) = \mathrm{Com}_{V_{A_{26}}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) = \bigoplus_{\lambda\in\frac{1}{9}Y/Y} V_{\lambda+Y} \otimes K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 9\bar{\lambda}).$$

Take $i = 3k$ and $j = 3\ell$. Then we have

$$\begin{aligned} \mathrm{Com}_{V_{(k\nu_1+\ell\nu_2)(\delta)+A_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) &\cong K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 9 \cdot -\tfrac{1}{9}(3k\alpha_1 + 3\ell\alpha_2)) \\ &= K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, -3(k\alpha_1 + \ell\alpha_2)) \end{aligned}$$

as desired. □

Lemma 4.10. *We have the decomposition*

$$\widetilde{W} = \mathrm{Com}_{V_{E_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) = \bigoplus_{i,j=0,\pm1} K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 3(i\alpha_1 + j\alpha_2)).$$

Proof. First we note that $M + N \cong A_2 \otimes E_8$ and

$$M + N = \bigcup_{0\le k,\ell\le 2} ((k\nu_1 + \ell\nu_2)(\delta) + A_2 \otimes A_8).$$

Since $A_2 \otimes A_8 \cong \mathrm{Ann}_{A_8^3}(d_3(A_8))$ and $V_{d_3(A_8)} \subset L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$, we have

$$\widetilde{W} = \mathrm{Com}_{V_{E_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) < V_{M+N}.$$

The conclusion now follows from Lemma 4.9. □

Now let $\alpha \in A_2$ be a root. Then $\mathbb{Z}\alpha \cong A_1$ and

$$L(\alpha) = \bigoplus_{j=1}^{9} \iota_j(\mathbb{Z}\alpha) \cong A_1^9 \subset A_{26}.$$

Let H_α and E_α be defined as in Notation 3.2. Then $\{H_\alpha, E_\alpha, -E_{-\alpha}\}$ forms a $\mathfrak{sl}_2$-triple in the lattice VOA $V_{A_1^9} < V_{A_{26}}$. Moreover, it generates a subVOA $\mathscr{L}_\alpha$ isomorphic to the affine VOA $L_{\widehat{\mathfrak{sl}}_2(\mathbb{C})}(9,0)$. Let $M_\alpha(9,0)$ be the subVOA generated by H_α. Then

$$\mathscr{K}_\alpha := \mathrm{Com}_{\mathscr{L}_\alpha}(M_\alpha(9,0)) \cong K(\mathfrak{sl}_2(\mathbb{C}), 9).$$

Note also that $\mathscr{K}_\alpha = \mathrm{Com}_{\mathscr{L}_\alpha}(M_\alpha(9,0)) < \mathrm{Com}_{L_{\widehat{\mathfrak{sl}}_3(\mathbb{C})}(9,0)}(V_{\mu(A_2)}) = K(\mathfrak{sl}_3(\mathbb{C}), 9)$.

Set $h_\alpha = H_\alpha$, $x_\alpha = E_\alpha$ and $x_{-\alpha} = -E_{-\alpha}$. Then the elements ω_α and W_α^3 defined in Notation 4.3 are contained in $\mathscr{K}_\alpha$. In fact, $\mathscr{K}_\alpha$ is generated by ω_α and W_α^3 (see [Dong et al. 2009]).

Theorem 4.11. *The VOA W defined in Notation 3.8 contains a full subVOA isomorphic to $K(\mathfrak{sl}_3(\mathbb{C}), 9)$.*

Proof. Recall that $W = \langle e^{i,j} \mid 0 \le i, j \le 2\rangle$. We also have

$$M = (\eta_1 - \eta_2)(E_8), \quad N = (\eta_2 - \eta_3)(E_8), \quad \widetilde{N} = (\eta_1 - \eta_3)(E_8).$$

Let $\alpha_1 = (1,-1,0)$, $\alpha_2 = (0,1,-1)$ and $\alpha_3 = \alpha_1 + \alpha_2 = (1,0,-1)$ be the positive roots of A_2. Then by the same calculations as in [Lam et al. 2007], it is straightforward to verify that

$$\mathscr{K}_{\alpha_1} < \mathrm{VOA}(e_M, \rho e_M), \quad \mathscr{K}_{\alpha_2} < \mathrm{VOA}(e_N, \rho e_N), \quad \mathscr{K}_{\alpha_3} < \mathrm{VOA}(e_{\widetilde{N}}, \rho e_{\widetilde{N}}),$$

where $e_M, e_N, e_{\widetilde{N}}$ and ρ are defined as in Notation 3.6.

Now by Theorem 4.4, $\mathscr{K}_{\alpha_1}, \mathscr{K}_{\alpha_2}$ and $\mathscr{K}_{\alpha_3}$ generate a subVOA isomorphic to $K(\mathfrak{sl}_3(\mathbb{C}), 9)$ in W. It is a full subVOA of W because they have the same central charge. □

Theorem 4.12. *We have $W = \widetilde{W} = \mathrm{Com}_{V_{E_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0))$.*

Proof. By the previous lemma, the subVOA W contains $K(\mathfrak{sl}_3(\mathbb{C}), 9)$ as a full subVOA.

Therefore, it suffices to show that $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 3(i\alpha_1 + j\alpha_2))$ is contained in W for any $i, j = 0, \pm 1$.

By [Lam et al. 2007, Proposition 2.2],

$$X^+_{\nu_1} = \frac{1}{32} \sum_{\substack{\gamma\in\nu_1(\delta)+\nu_1(A_8)\\ \langle\gamma,\gamma\rangle=4}} e^\gamma \quad\text{and}\quad X^-_{\nu_1} = \frac{1}{32} \sum_{\substack{\gamma\in-\nu_1(\delta)+\nu_1(A_8)\\ \langle\gamma,\gamma\rangle=4}} e^\gamma$$

are contained in $\mathrm{VOA}(e_M, \rho e_M) < W$. Moreover, it is straightforward to verify that

$$X^+_{\nu_1} \in \mathrm{Com}_{V_{\nu_1(\delta)+A_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) \cong K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,-3\alpha_1)$$

and

$$X^-_{\nu_1} \in \mathrm{Com}_{V_{-\nu_1(\delta)+A_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)) \cong K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,3\alpha_1).$$

Therefore, W contains $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,\pm 3\alpha_1)$ as $K(\mathfrak{sl}_3(\mathbb{C}),9)$-submodules. Similarly, W also contains $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,\pm 3\alpha_2)$ and $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,\pm 3(\alpha_1+\alpha_2))$ as $K(\mathfrak{sl}_3(\mathbb{C}),9)$-submodules.

Moreover, it is clear that $0 \neq (X^+_{\nu_1})_{-3}(X^-_{\nu_2}) \in V_{(\nu_1-\nu_2)(\delta)+A_8^3}$. Since $X^+_{\nu_1}$ and $X^-_{\nu_2}$ are contained in the commutant of $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)$, we have

$$(X^+_{\nu_1})_{-3}(X^-_{\nu_2}) \in \mathrm{Com}_{V_{(\nu_1-\nu_2)(\delta)+A_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3,0)).$$

Hence W contains $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,3(\alpha_1-\alpha_2))$. Similarly, $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,3(\alpha_2-\alpha_1))$ is contained in W, also. $\square$

Remark 4.13. Recall that $e^{i,j} \in V_{E_8^3}$, $0 \le i, j \le 2$, are fixed by the diagonal action of the Weyl group of K. Therefore, the VOA $\widetilde{W}$ is fixed by the Weyl group of K pointwise. Using this fact and Lemma 4.9, it is straightforward to show that $\dim(K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,3\alpha)_2) = 1$ for any root α of A_8, $\dim(K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)_2) = 3$, and $\dim(K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,\pm 3(\alpha_1-\alpha_2))_2) = 0$ (see the Appendix). Thus, $\dim(\widetilde{W}_2) = 9$ and $\widetilde{W}_2$ is spanned by $\{e^{i,j} \mid 0 \le i, j \le 2\}$.

A positive definite real form. Next we shall show that the Ising vectors $e^{i,j}$, for $0 \le i, j \le 2$, are contained in a positive definite real form of $V_{E_8^3}$.

First we recall that the lattice VOA constructed in [Frenkel et al. 1988] can be defined over $\mathbb{R}$. Let $V_{L,\mathbb{R}} = S(\hat{\mathfrak{h}}^-_{\mathbb{R}}) \otimes \mathbb{R}\{L\}$ be the real lattice VOA associated to an even positive definite lattice, where $\mathfrak{h} = \mathbb{R} \otimes_{\mathbb{Z}} L$, $\hat{\mathfrak{h}}^- = \oplus_{n\in\mathbb{Z}^+} \mathfrak{h} \otimes \mathbb{R}t^{-n}$. As usual, we use $x(-n)$ to denote $x \otimes t^{-n}$ for $x \in \mathfrak{h}$ and $n \in \mathbb{Z}^+$.

Notation 4.14. Let $\theta : V_{L,\mathbb{R}} \to V_{L,\mathbb{R}}$ be defined by

$$\theta(x(-n_1)\cdots x(-n_k) \otimes e^\alpha) = (-1)^k x(-n_1)\cdots x(-n_k) \otimes e^{-\alpha}.$$

Then θ is an automorphism of $V_{L,\mathbb{R}}$, which is a lift of the (-1)-isometry of L [ibid.]. We shall denote the (± 1)-eigenspaces of θ on $V_{L,\mathbb{R}}$ by $V^\pm_{L,\mathbb{R}}$.

The following result is well-known [Frenkel et al. 1988; Miyamoto 2004].

Proposition 4.15 (cf. Proposition 2.7 of [Miyamoto 2004]). *Let L be an even positive definite lattice. Then the real subspace $\widetilde{V}_{L,\mathbb{R}} = V^+_{L,\mathbb{R}} \oplus \sqrt{-1}V^-_{L,\mathbb{R}}$ is a real form of V_L. Furthermore, the invariant form on $\widetilde{V}_{L,\mathbb{R}}$ is positive definite.*

Now apply the above theorem to the case $L = E_8^3$. We have the following result.

Proposition 4.16. *Let $\widetilde{V}_{E_8^3,\mathbb{R}} = V^+_{E_8^3,\mathbb{R}} \oplus \sqrt{-1}V^-_{E_8^3,\mathbb{R}}$. Then $\widetilde{V}_{E_8^3,\mathbb{R}}$ is a positive definite real form of $V_{E_8^3}$.*

The next lemma is clear by the definitions of e_N, e_N, and $e_{\widetilde{N}}$.

Lemma 4.17. *The Ising vectors e_M, e_N and $e_{\widetilde{N}}$ defined in Notation 3.6 lie in $V^+_{E_8^3,\mathbb{R}}$.*

Recall the automorphism $\rho = \exp\big(\frac{2\pi i}{3}(\boldsymbol{a}, -\boldsymbol{a}, 0)(0)\big)$ defined in Definition 3.5, where $\boldsymbol{a}$ is an element of E_8 such that $K = \{\beta \in E_8 \mid \langle \beta, \boldsymbol{a}\rangle \in 3\mathbb{Z}\} \cong A_8$. Then we have the coset decomposition

$$E_8 = A_8 \cup (b + A_8) \cup (-b + A_8),$$

where b is a root of E_8 such that $\langle b, \boldsymbol{a}\rangle \equiv 1 \bmod 3$.

Note that

$$\begin{aligned} M &= \{(\alpha, -\alpha, 0) \mid \alpha \in E_8\} \cong \sqrt{2}E_8, \\ \widetilde{K} &= \{(\alpha, -\alpha, 0) \mid \alpha \in K\} \cong \sqrt{2}A_8. \end{aligned}$$

Set

$$\begin{aligned} X^0 &:= \tfrac{1}{3}(e_M + \rho e_M + \rho^2 e_M), \\ X^1 &:= \tfrac{1}{3}(e_M + \xi\rho e_M + \xi^2\rho^2 e_M), \\ X^2 &:= \tfrac{1}{3}(e_M + \xi^2\rho e_M + \xi\rho^2 e_M), \end{aligned}$$

where $\xi = \exp\frac{2\pi i}{3} = \frac{1}{2}(-1 + \sqrt{-3})$.

The next lemma can be proved by the same calculations as in [Lam et al. 2007]. Note that $\rho X^0 = X^0$, $\rho X^1 = \xi^2 X^1$ and $\rho X^2 = \xi X^2$.

Lemma 4.18. *The vector X^0 is contained in $V^+_{M,\mathbb{R}}$. Moreover,*

$$X^1 = \frac{1}{32} \sum_{\substack{\gamma \in (b,-b,0)+\widetilde{K} \\ \langle \gamma,\gamma\rangle = 4}} e^\gamma \quad \textit{and} \quad X^2 = \frac{1}{32} \sum_{\substack{\gamma \in -(b,-b,0)+\widetilde{K} \\ \langle \gamma,\gamma\rangle = 4}} e^\gamma .$$

Therefore, $X^1 + X^2 \in V^+_{M,\mathbb{R}}$ and $X^1 - X^2 \in V^-_{M,\mathbb{R}}$.

Lemma 4.19. *The Ising vectors $e^{i,j}$, $0 \le i, j \le 2$, are all contained in $\widetilde{V}_{E_8^3,\mathbb{R}}$.*

Proof. By the discussion above, we have

$$\rho e_M = X^0 - \tfrac{1}{2}(X^1 + X^2) + \tfrac{1}{2}\sqrt{-3}(X^1 - X^2).$$

Since $X^1 + X^2 \in V^+_{M,\mathbb{R}}$ and $X^1 - X^2 \in V^-_{M,\mathbb{R}}$, we have $\rho e_M \in \widetilde{V}_{E_8^3,\mathbb{R}}$. Similarly, we have $\rho^2 e_M$, ρe_N, $\rho^2 e_N$, $\rho e_{\widetilde{N}}$, $\rho^2 e_{\widetilde{N}} \in \widetilde{V}_{E_8^3,\mathbb{R}}$ as desired. □

5. Griess algebras generated by Ising vectors

In this section, we shall give few structural results about Griess algebras generated by Ising vectors in a moonshine-type VOA V over $\mathbb{R}$ such that the invariant bilinear form is positive definite. Our setting is as follows.

Notation 5.1. Let e, e', e'' be three distinct Ising vectors in V. Assume that

(I) $\langle e, e'\rangle = \langle e, e'\rangle = \langle e', e''\rangle = 1/2^8$ and $\tau_e\tau_{e'}$, $\tau_e\tau_{e''}$, $\tau_{e'}\tau_{e''}$ are of order 3.

Then each of $\{\tau_e, \tau_{e'}\}$, $\{\tau_e, \tau_{e''}\}$, and $\{\tau_{e'}, \tau_{e''}\}$ generates a dihedral group of order 6 and the Griess algebras generated by $\{e, e'\}$, $\{e, e''\}$, and $\{e', e''\}$ are isomorphic to the Griess algebra $\mathcal{G}U_{3C}$ of the $3C$-algebra U_{3C}.

Let $g = \tau_e\tau_{e'}$ and $h = \tau_e\tau_{e''}$. We shall assume that

(II) the subgroup H generated by g and h is elementary abelian of order 3^2.

For any $0 \le i, j \le 2$, denote $e^{i,j} := g^i h^j e$. Note that $e' = ge = e^{1,0}$ and $e'' = he = e^{0,1}$ by Lemma 2.9(4). Furthermore, we assume that

(III) $\langle e^{0,0}, e^{1,1}\rangle = 1/2^8$.

Therefore, the Griess subalgebra generated by $e^{0,0}$, $e^{1,1}$ is also isomorphic to $\mathcal{G}U_{3C}$.

Lemma 5.2. *Let G be the subgroup generated by τ_e, $\tau_{e'}$, and $\tau_{e''}$. Then $G = H : \langle\tau_e\rangle$, where $H \cong 3^2$ is normal in G and τ_e inverts every element in H, i.e., $\tau_e y \tau_e = y^{-1}$ for all $y \in H$.*

Proof. The proof is essentially the same as Lemma 3.12 because $H = \langle g, h\rangle$ is elementary abelian of order 3^2 by our assumption. □

Lemma 5.3. *For any $(i, j) \neq (i', j')$, we have $\langle e^{i,j}, e^{i',j'}\rangle = 1/2^8$.*

Proof. By definition, $\langle e^{i,j}, e^{i',j'}\rangle = \langle g^i h^j e, g^{i'} h^{j'} e\rangle = \langle e, g^{i'-i} h^{j'-j} e\rangle$.

By our assumption, we have

$$\begin{aligned} \langle e, ge\rangle = \langle e, g^{-1}e\rangle = \langle e, he\rangle = \langle e, h^{-1}e\rangle &= 1/2^8, \\ \langle e, gh^{-1}e\rangle = \langle e, g^{-1}he\rangle = \langle ge, he\rangle = \langle e', e''\rangle &= 1/2^8, \\ \langle e, ghe\rangle = \langle e, g^{-1}h^{-1}e\rangle &= 1/2^8. \end{aligned}$$

Thus, $\langle e^{i,j}, e^{i',j'}\rangle = 1/2^8$ if $(i, j) \neq (i', j')$. □

Lemma 5.4. *Let $\mathcal{G}$ be the Griess subalgebra generated by $\{e, e', e''\}$. Then $\mathcal{G}$ is spanned by $\{e^{i,j} \mid 0 \le i, j \le 2\}$ and $\dim \mathcal{G} = 9$. The algebra structure of $\mathcal{G}$ is unique.*

Proof. Recall that g commutes with h and for any (i, j) and (i', j'), we have

$$\begin{aligned}\tau_{e^{i,j}}\tau_{e^{i',j'}} &= g^i h^j \tau_e g^{-i} h^{-j} g^{i'} h^{j'} \tau_e g^{-i'} h^{-j'} = g^i h^j g^i h^j \tau_e \tau_e g^{-i'} h^{-j'} g^{-i'} h^{-j'} \\ &= g^{i'-i} h^{j'-j}.\end{aligned}$$

By Lemma 2.9(3) and (4), we know that

$$e^{-i-i',-j-j'} = g^{i'-i} h^{j'-j}(e^{i,j}) = e^{i,j} + e^{i',j'} - 32 e^{i,j} \cdot e^{i',j'} \tag{5-1}$$

if $(i, j) \neq (i'j')$. Therefore, $e^{-i-i',-j-j'} \in \mathcal{G}\{e^{i,j}, e^{i',j'}\}$, the Griess subalgebra generated by $\{e^{i,j}, e^{i',j'}\}$. Hence,

$$e^{2,0} = e^{0,0} + e^{1,0} - 32 e^{0,0} \cdot e^{1,0},\ e^{0,2} = e^{0,0} + e^{0,1} - 32 e^{0,0} \cdot e^{0,1},$$

and $e^{2,2} = e^{1,0} + e^{0,1} - 32 e^{1,0} \cdot e^{0,1}$ are in $\mathcal{G}$.

Similarly, we also have $e^{1,1} \in \mathcal{G}\{e^{0,0}, e^{2,2}\} < \mathcal{G}$, $e^{1,2} \in \mathcal{G}\{e^{0,2}, e^{2,2}\} < \mathcal{G}$, and $e^{2,1} \in \mathcal{G}\{e^{2,0}, e^{2,2}\} < \mathcal{G}$. Thus, all $e^{i,j}$, $0 \le i, j \le 2$, are in $\mathcal{G}$. In addition, by (5-1), we have

$$e^{i,j} \cdot e^{i',j'} = \begin{cases} \frac{1}{32}\left(e^{i,j} + e^{i',j'} - e^{i'',j''}\right) & \text{if } (i, j) \neq (i'j'), \\ 2e^{i,j} & \text{if } (i, j) = (i'j'), \end{cases}$$

where $i + i' + i'' = j + j' + j'' = 0 \bmod 3$. Therefore, $\operatorname{span}\{e^{i,j} \mid 0 \le i, j \le 2\}$ is closed under the Griess algebra product and $\mathcal{G} = \operatorname{span}\{e^{i,j} \mid 0 \le i, j \le 2\}$. By our assumption, we have the Gram matrix

$$\left(\langle e^{i,j}, e^{i',j'}\rangle\right)_{0 \le i,j,i',j' \le 2} = \begin{pmatrix} \frac{1}{4} & \frac{1}{2^8} & \cdots & \frac{1}{2^8} \\ \frac{1}{2^8} & \frac{1}{4} & \cdots & \frac{1}{2^8} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2^8} & \frac{1}{2^8} & \cdots & \frac{1}{4} \end{pmatrix}.$$

It has rank 9 and hence $\{e^{i,j} \mid 0 \le i, j \le 2\}$ is a linearly independent set and $\dim \mathcal{G} = 9$. □

Next, we shall give some information about the VOA W generated by $\{e^{i,j}\}$.

Lemma 5.5. *Let*

$$\omega := \frac{8}{9} \sum_{0 \le i,j \le 2} e^{i,j}.$$

Then ω is a Virasoro vector of central charge 4. *Moreover, $\omega \cdot e^{i,j} = e^{i,j} \cdot \omega = 2e^{i,j}$ for any $0 \le i, j \le 2$. In other words, $\omega/2$ is the identity element in $\mathcal{G}$.*

Proof. This follows from a straightforward calculation using Lemma 2.9. □

Lemma 5.6. *Let*

$$b^1 = \tfrac{8}{9} \sum_{0 \le i,j \le 2} e^{i,j} - \tfrac{32}{33}(e^{0,0} + e^{0,1} + e^{0,2}).$$

Then b^1 is a Virasoro vector of central charge $\frac{28}{11}$. Moreover, $e^{0,0}$, a^1, and b^1 are mutually orthogonal and $\omega = e^{0,0} + a^1 + b^1$. Therefore, W has a full subVOA isomorphic to the tensor product of Virasoro VOA

$$L\big(\tfrac{1}{2}, 0\big) \otimes L\big(\tfrac{21}{22}, 0\big) \otimes L\big(\tfrac{28}{11}, 0\big).$$

Proof. It follows from (4) and (5) of Lemma 2.9 and Lemma 5.5. □

Remark 5.7. Because of Lemma 4.10 and Theorem 4.12, we conjecture that the subVOA VOA(e, e', e'') generated by $\{e, e', e''\}$ is isomorphic to

$$\widetilde{W} = \bigoplus_{i,j=0,\pm 1} K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 3(i\alpha_1 + j\alpha_2)).$$

Recall from [Lam 2014] that the parafermion VOA $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)$ contains a full subVOA $W_9(1,1) \otimes W_9(2,1)$, where $W_9(1,1)$ has central charge $\frac{32}{11}$ and $W_9(2,1)$ has central charge $\frac{28}{11}$. Therefore, we believe that the subVOA VOA(e, e', e'') also contains a full subVOA isomorphic to $L(\frac{1}{2}, 0) \otimes L(\frac{21}{22}, 0) \otimes W_9(2,1)$, which is expected to be rational. However, we are not aware of any uniqueness results of the parafermion VOA $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)$ nor the W-algebra $W_9(2,1)$ in terms of generators and relations. Therefore, it is unclear if VOA(e, e', e'') contains $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)$ or $L(\frac{1}{2}, 0) \otimes L(\frac{21}{22}, 0) \otimes W_9(2,1)$ as a full subVOA.

Finally, we describe explicitly several highest weight vectors of the subVOA $\mathrm{vir}(e^{0,0}) \otimes \mathrm{vir}(a^1) \otimes \mathrm{vir}(b^1)$.

Lemma 5.8. *With respect to the subVOA* $\mathrm{vir}(e^{0,0}) \otimes \mathrm{vir}(a^1) \otimes \mathrm{vir}(b^1)$, *we have the following highest weight vectors.*

(1) *The vectors $a^i - a^j$, $i, j \in \{2,3,4\}$, $i \neq j$, are highest weight vectors of weight $(0, \frac{1}{11}, \frac{21}{11})$, where*

$$a^1 = \tfrac{32}{33}(e^{0,0} + e^{0,1} + e^{0,2}) - e^{0,0}, \quad a^2 = \tfrac{32}{33}(e^{0,0} + e^{1,0} + e^{2,0}) - e^{0,0},$$
$$a^3 = \tfrac{32}{33}(e^{0,0} + e^{1,1} + e^{2,2}) - e^{0,0}, \quad a^4 = \tfrac{32}{33}(e^{0,0} + e^{1,2} + e^{2,1}) - e^{0,0}.$$

(2) *The vector $e^{0,1} - e^{0,2}$ is a highest weight vector of weight $(\frac{1}{16}, \frac{31}{16}, 0)$.*

(3) *The vector $(e^{1,0} + e^{1,1} + e^{1,2}) - (e^{2,0} + e^{2,1} + e^{2,2})$ is a highest weight vector of weight $(\frac{1}{16}, \frac{21}{176}, \frac{20}{11})$.*

(4) *The vectors $(e^{1,1} - e^{2,2}) - (e^{1,2} - e^{2,1})$ and $(e^{1,0} - e^{2,0}) - (e^{1,1} - e^{2,2})$ are highest weight vectors of weight $(\frac{1}{16}, \frac{5}{176}, \frac{21}{11})$.*

Proof. (1) By Lemma 2.9, it is straightforward to show that

$$a^i \cdot a^j = \tfrac{1}{33}(2a^i + 2a^j - a^k - a^\ell),$$

for any $i \neq j$ and $\{i, j, k, \ell\} = \{1, 2, 3, 4\}$. Thus,

$$a^1_1\big(a^i - a^j\big) = \tfrac{1}{33}\big[\big(2a^1 + 2a^i - a^j - a^k\big) - \big(2a^1 + 2a^j - a^i - a^k\big)\big] = \tfrac{1}{11}\big(a^i - a^j\big),$$

where $\{i, j, k\} = \{2, 3, 4\}$.

Since $e^{0,0}_1 a^i = 0$ and $\omega_1 a^i = 2a^i$ for all i, $a^i - a^j$ is a highest weight vector of weight $(0, \frac{1}{11}, \frac{21}{11})$ with respect to $\mathrm{vir}(e^{0,0}) \otimes \mathrm{vir}(a^1) \otimes \mathrm{vir}(b^1)$.

(2) By direct calculations, we have

$$e^{0,0}_1\big(e^{0,1} - e^{0,2}\big) = \tfrac{1}{32}\big[\big(e^{0,0} + e^{0,1} - e^{0,2}\big) - \big(e^{0,0} + e^{0,2} - e^{0,1}\big)\big] = \tfrac{1}{16}\big(e^{0,1} - e^{0,2}\big)$$

and

$$\tfrac{32}{33}\big(e^{0,0} + e^{0,1} + e^{0,2}\big)_1\big(e^{0,1} - e^{0,2}\big) = 2\big(e^{0,1} - e^{0,2}\big).$$

Since $a^1 = \frac{32}{33}(e^{0,0} + e^{0,1} + e^{0,2}) - e^{0,0}$ and $b^1 = \omega - \frac{32}{33}(e^{0,0} + e^{0,1} + e^{0,2})$, we have

$$a^1_1\big(e^{0,1} - e^{0,2}\big) = \tfrac{31}{16}\big(e^{0,1} - e^{0,2}\big) \quad \text{and} \quad b^1_1\big(e^{0,1} - e^{0,2}\big) = 0.$$

(3), (4) By the same calculations as in (2), $(e^{1,0} - e^{2,0})$, $(e^{1,1} - e^{2,2})$, and $(e^{1,2} - e^{2,1})$ are $\frac{1}{16}$-eigenvectors of $e^{0,0}_1$. By Lemma 2.9, we also have

$$\tfrac{32}{33}\big(e^{0,0} + e^{0,1} + e^{0,2}\big)_1\big(e^{1,1} - e^{2,2}\big) = \tfrac{1}{33}\big(4\big(e^{1,1} - e^{2,2}\big) + \big(e^{1,0} - e^{2,0}\big) + \big(e^{1,2} - e^{2,1}\big)\big).$$

Let $v = (e^{1,0} + e^{1,1} + e^{1,2}) - (e^{2,0} + e^{2,1} + e^{2,2})$. Then

$$\tfrac{32}{33}\big(e^{0,0} + e^{0,1} + e^{0,2}\big)_1 v = \tfrac{1}{33}(4 + 1 + 1)v = \tfrac{2}{11}v.$$

Thus, $a^1_1 v = (\frac{2}{11} - \frac{1}{16})v = \frac{21}{176}v$ and $b^1_1 v = (2 - \frac{2}{11})v = \frac{20}{11}v$.

Moreover,

$$\begin{aligned}
&\tfrac{32}{33}\big(e^{0,0} + e^{0,1} + e^{0,2}\big)_1\big(\big(e^{1,1} - e^{2,2}\big) - \big(e^{1,2} - e^{2,1}\big)\big) \\
&\qquad = \tfrac{1}{33}(4 - 1)\big(\big(e^{1,1} - e^{2,2}\big) - \big(e^{1,2} - e^{2,1}\big)\big) = \tfrac{1}{11}\big(\big(e^{1,1} - e^{2,2}\big) - \big(e^{1,2} - e^{2,1}\big)\big).
\end{aligned}$$

Thus, we have

$$a^1_1\big(\big(e^{1,1} - e^{2,2}\big) - \big(e^{1,2} - e^{2,1}\big)\big) = \tfrac{5}{176}\big(\big(e^{1,1} - e^{2,2}\big) - \big(e^{1,2} - e^{2,1}\big)\big)$$

and

$$b^1_1\big(\big(e^{1,1} - e^{2,2}\big) - \big(e^{1,2} - e^{2,1}\big)\big)v = \tfrac{21}{11}\big(\big(e^{1,1} - e^{2,2}\big) - \big(e^{1,2} - e^{2,1}\big)\big).$$

The remaining cases can be proved similarly. □

Appendix: Dimensions of $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 3(i\alpha_1 + j\alpha_2))_2$

In this appendix, we shall compute the dimension of $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 3(i\alpha_1 + j\alpha_2))_2$ for all $0 \le i, j \le 2$. First we recall a result from [Frenkel et al. 1988, Chapter 8].

Let α, β have norm 4 in a lattice L. Then

$$e_1^{\alpha} e^{\beta} = \begin{cases} \frac{1}{2}\alpha(-1)^2 \cdot \mathbb{1} & \text{if } \beta = -\alpha, \\ \varepsilon(\alpha, \beta) e^{\alpha+\beta} & \text{if } \langle \beta, \alpha \rangle = -2, \\ 0 & \text{otherwise.} \end{cases} \tag{A-1}$$

Lemma A.1. *For any* $k, \ell = 0, \pm 1$, *we have* $\dim(K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 3(k\alpha_1 + \ell\alpha_2))_2) = 1$ *if* $(k, \ell) = (0, \pm 1)$, $(\pm 1, 0)$, *or* $(\pm 1, \pm 1)$ *and* $\dim(K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 3(k\alpha_1 + \ell\alpha_2))_2) = 0$ *if* $(k, \ell) = (\pm 1, \mp 1)$.

Proof. Recall that

$$K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, -3(k\alpha_1 + \ell\alpha_2)) \cong \mathrm{Com}_{V_{(k\nu_1+\ell\nu_2)(\delta)+A_8^3}}(L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3, 0)) < V_{M+N}.$$

Moreover, by Lemma 3.3, the conformal vector Ω of $L_{\widehat{\mathfrak{sl}}_9(\mathbb{C})}(3, 0)$ is given by

$$\Omega = \omega_E + \tfrac{3}{4}\omega_{M+N} - \tfrac{1}{12} \sum_{\substack{\alpha \in A_8(2) \\ 1 \le i, j \le 3,\, i \ne j}} e^{\eta_i(\alpha) - \eta_j(\alpha)}.$$

Thus, for any $X \in (V_{M+N})_2$, $\Omega_1 X = 0$ if and only if

$$\Bigg(\sum_{\substack{\alpha \in A_8(2) \\ 1 \le i, j \le 3,\, i \ne j}} e^{\eta_i(\alpha) - \eta_j(\alpha)} \Bigg)_1 X = 18X. \tag{A-2}$$

Let $\Psi = \{\gamma \in (k\nu_1 + \ell\nu_2)(\delta) + A_8^3 \mid \langle \gamma, \gamma \rangle = 4\}$. Then

$$(V_{(k\nu_1+\ell\nu_2)(\delta)+A_8^3})_2 = \mathrm{span}\{e^{\gamma} \mid \gamma \in \Psi\}.$$

Moreover, $\Psi = \{(k\nu_1 + \ell\nu_2)(\delta + \beta) \mid \beta \in A_8(2), \langle \beta, \delta \rangle = -1\}$ if $(k, \ell) = (0, \pm 1)$, $(\pm 1, 0)$, or $(\pm 1, \pm 1)$ and $\Psi = \varnothing$ if $(k, \ell) = (\pm 1, \mp 1)$.

Suppose $X = \sum_{\gamma \in \Psi} a_\gamma e^{\gamma}$ be an element in $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, -3(k\alpha_1 + \ell\alpha_2))_2$. Then by (A-1) and (A-2), we must have $a_\gamma = a_{\gamma'}$ for all $\gamma, \gamma' \in \Psi$. Note that there are exactly 18 roots in A_8 such that $\langle \delta, \beta \rangle = -1$ and $\widetilde{W}$ is fixed pointwise by the Weyl group of $K \cong A_8$.

Hence, $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, -3(k\alpha_1 + \ell\alpha_2))_2$ is spanned by $\sum_{\gamma \in \Psi} e^{\gamma}$ if $(k, \ell) = (0, \pm 1)$, $(\pm 1, 0)$, or $(\pm 1, \pm 1)$ and is zero if $(k, \ell) = (\pm 1, \mp 1)$. □

Next we consider the Griess algebra $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0, 0)_2$. The next lemma follows immediately from (A-1) and the choice of the 2-cocycle $\varepsilon_0(\ ,\ )$.

Lemma A.2. *Let β be a root of A_8 and $1 \le k, \ell \le 3$. Then*

$$\Bigg(\sum_{\substack{\alpha\in A_8(2)\\ i\neq j}} e^{(\eta_i-\eta_j)(\alpha)}\Bigg)_1 e^{(\eta_k-\eta_\ell)(\beta)}$$
$$= \sum_{\substack{\alpha\in A_8(2)\\ \langle\alpha,\beta\rangle=-1}} e^{(\eta_k-\eta_\ell)(\alpha+\beta)} + \tfrac{1}{2}(\eta_k(\beta)-\eta_\ell(\beta))(-1)^2\cdot\mathbb{1} - \sum_{i\neq k} e^{(\eta_i-\eta_\ell)(\beta)} - \sum_{j\neq\ell} e^{(\eta_k-\eta_j)(\beta)}.$$

The next lemma can also be proved easily by the definition of vertex operators [Frenkel et al. 1988].

Lemma A.3. *Let $\beta \in A_8$. Then*

$$\Bigg(\sum_{\substack{\alpha\in A_8(2)\\ i\neq j}} e^{(\eta_i-\eta_j)(\alpha)}\Bigg)_1 (\eta_k-\eta_\ell)(\beta))(-1)^2\cdot\mathbb{1}$$
$$= \sum_{\substack{\alpha\in A_8(2)\\ i\neq j}} \langle(\eta_i-\eta_j)(\alpha), (\eta_k-\eta_\ell)(\beta)\rangle^2 e^{(\eta_i-\eta_j)(\alpha)}.$$

Lemma A.4. *The Griess algebra $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)_2$ has dimension 3 and is spanned by $\{\omega_{\alpha_1}, \omega_{\alpha_2}, \omega_{\alpha_1+\alpha_2}\}$.*

Proof. Let

$$X = \sum_{\substack{1\le i<j\le 3,\\ \alpha\in A_8(2)}} \big(a_{i,j,\alpha}(\eta_i(\alpha)-\eta_j(\alpha))(-2)\cdot\mathbb{1} + b_{i,j,\alpha}(\eta_i(\alpha)-\eta_j(\alpha))(-1)^2\cdot\mathbb{1} + c_{i,j,\alpha}e^{\eta_i(\alpha)-\eta_j(\alpha)}\big)$$

be an element in $K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)_2$.

Since X is fixed by the Weyl group of A_8, we have $a_{i,j,\alpha} = a_{i,j,\beta}$, $b_{i,j,\alpha} = b_{i,j,\beta}$, and $c_{i,j,\alpha} = c_{i,j,\beta}$ for any roots $\alpha, \beta \in A_8$. Set $a_{i,j} = a_{i,j,\alpha}$, $b_{i,j} = b_{i,j,\alpha}$, and $c_{i,j} = c_{i,j,\alpha}$ for any root $\alpha \in A_8$. Then, for any $1 \le i < j \le 3$,

$$\sum_{\alpha\in A_8(2)} a_{i,j,\alpha}(\eta_i(\alpha)-\eta_j(\alpha))(-2) = a_{i,j}\sum_{\alpha\in A_8(2)}(\eta_i(\alpha)-\eta_j(\alpha))(-2) = 0$$

and

$$X = \sum_{1\le i<j\le 3}\Bigg(b_{i,j}\sum_{\alpha\in A_8(2)}(\eta_i(\alpha)-\eta_j(\alpha))(-1)^2\cdot\mathbb{1} + c_{i,j}\sum_{\alpha\in A_8(2)} e^{\eta_i(\alpha)-\eta_j(\alpha)}\Bigg).$$

Moreover, $\Bigg(\sum_{\substack{\alpha\in A_8(2)\\ 1\le i,j\le 3,\, i\neq j}} e^{\eta_i(\alpha)-\eta_j(\alpha)}\Bigg)_1 X = 18X$ since $X \in K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)_2$.

By Lemmas A.2 and A.3, it is straightforward to show $X \in \mathrm{span}\{\omega_{\alpha_1}, \omega_{\alpha_2}, \omega_{\alpha_1+\alpha_2}\}$ and $\dim(K_{\mathfrak{sl}_3(\mathbb{C}),9}(0,0)_2) = 3$. $\square$

References

[Conway et al. 1985] J. H. Conway, R. T. Curtis, S. P. Norton, R. A. Parker, and R. A. Wilson, *Atlas of finite groups: maximal subgroups and ordinary characters for simple groups*, Oxford University Press, Eynsham, 1985. MR 88g:20025 Zbl 0568.20001

[Dong and Lepowsky 1993] C. Dong and J. Lepowsky, *Generalized vertex algebras and relative vertex operators*, Progress in Mathematics **112**, Birkhäuser, Boston, 1993. MR 95b:17032 Zbl 0803.17009

[Dong and Wang 2010] C. Dong and Q. Wang, "The structure of parafermion vertex operator algebras: general case", *Comm. Math. Phys.* **299**:3 (2010), 783–792. MR 2011h:17037 Zbl 1239.17021

[Dong et al. 1994] C. Dong, G. Mason, and Y. Zhu, "Discrete series of the Virasoro algebra and the moonshine module", pp. 295–316 in *Algebraic groups and their generalizations: quantum and infinite-dimensional methods* (University Park, PA, 1991), edited by W. J. Haboush and B. J. Parshall, Proc. Sympos. Pure Math. **56** Part 2, American Mathematical Society, Providence, RI, 1994. MR 95c:17043 Zbl 0813.17019

[Dong et al. 1998] C. Dong, H. Li, G. Mason, and S. P. Norton, "Associative subalgebras of the Griess algebra and related topics", pp. 27–42 in *The Monster and Lie algebras* (Columbus, OH, 1996), edited by J. Ferrar and K. Harada, Ohio State Univ. Math. Res. Inst. Publ. **7**, De Gruyter, Berlin, 1998. MR 99k:17048 Zbl 0946.17011 arXiv q-alg/9607013

[Dong et al. 2009] C. Dong, C. H. Lam, and H. Yamada, "W-algebras related to parafermion algebras", *J. Algebra* **322**:7 (2009), 2366–2403. MR 2011b:17053 Zbl 1247.17019

[Dong et al. 2010] C. Dong, C. H. Lam, Q. Wang, and H. Yamada, "The structure of parafermion vertex operator algebras", *J. Algebra* **323**:2 (2010), 371–381. MR 2011a:17041 Zbl 1222.17021

[Frenkel and Zhu 1992] I. B. Frenkel and Y. Zhu, "Vertex operator algebras associated to representations of affine and Virasoro algebras", *Duke Math. J.* **66**:1 (1992), 123–168. MR 93g:17045 Zbl 0848.17032

[Frenkel et al. 1988] I. B. Frenkel, J. Lepowsky, and A. Meurman, *Vertex operator algebras and the Monster*, Pure and Applied Mathematics **134**, Academic Press, Boston, 1988. MR 90h:17026 Zbl 0674.17001

[Frenkel et al. 1993] I. B. Frenkel, Y.-Z. Huang, and J. Lepowsky, *On axiomatic approaches to vertex operator algebras and modules*, Memoirs of the American Mathematical Society **104**:494, American Mathematical Society, Providence, RI, 1993. MR 94a:17007 Zbl 0789.17022

[Griess 1982] R. L. Griess, Jr., "The friendly giant", *Invent. Math.* **69**:1 (1982), 1–102. MR 84m:20024 Zbl 0498.20013

[Griess and Lam 2011] R. L. Griess, Jr. and C. H. Lam, "EE_8-lattices and dihedral groups", *Pure Appl. Math. Q.* **7**:3 (2011), 621–743. MR 2012i:11068 Zbl 1247.11092 arXiv 0806.2753

[Höhn 2010] G. Höhn, "The group of symmetries of the shorter moonshine module", *Abh. Math. Semin. Univ. Hambg.* **80**:2 (2010), 275–283. MR 2012a:17052 Zbl 1210.17033 arXiv math/0210076

[Ivanov 2009] A. A. Ivanov, *The Monster group and Majorana involutions*, Cambridge Tracts in Mathematics **176**, Cambridge University Press, 2009. MR 2010h:20030 Zbl 1205.20014

[Ivanov 2011a] A. A. Ivanov, "Majorana representation of A_6 involving $3C$-algebras", *Bull. Math. Sci.* **1**:2 (2011), 365–378. MR 2901004 Zbl 1257.17036

[Ivanov 2011b] A. A. Ivanov, "On Majorana representations of A_6 and A_7", *Comm. Math. Phys.* **307**:1 (2011), 1–16. MR 2012h:20027 Zbl 1226.17023

[Ivanov and Seress 2012] A. A. Ivanov and Á. Seress, "Majorana representations of A_5", *Math. Z.* **272**:1-2 (2012), 269–295. MR 2968225 Zbl 1260.20019

[Ivanov et al. 2010] A. A. Ivanov, D. V. Pasechnik, Á. Seress, and S. Shpectorov, "Majorana representations of the symmetric group of degree 4", *J. Algebra* **324**:9 (2010), 2432–2463. MR 2011h:20026 Zbl 1257.20011

[Kac and Wakimoto 1988] V. G. Kac and M. Wakimoto, "Modular and conformal invariance constraints in representation theory of affine algebras", *Adv. in Math.* **70**:2 (1988), 156–236. MR 89h:17036 Zbl 0661.17016

[Lam 2014] C. H. Lam, "A level-rank duality for parafermion vertex operator algebras of type A", preprint, 2014. To appear in *Proc. Amer. Math. Soc.*

[Lam et al. 2005] C. H. Lam, H. Yamada, and H. Yamauchi, "McKay's observation and vertex operator algebras generated by two conformal vectors of central charge $1/2$", *Int. Math. Res. Pap.* **2005**:3 (2005), 117–181. MR 2006h:17034 Zbl 1082.17015 arXiv math/0503239

[Lam et al. 2007] C. H. Lam, H. Yamada, and H. Yamauchi, "Vertex operator algebras, extended E_8 diagram, and McKay's observation on the Monster simple group", *Trans. Amer. Math. Soc.* **359**:9 (2007), 4107–4123. MR 2008b:17046 Zbl 1139.17011

[Miyamoto 1996] M. Miyamoto, "Griess algebras and conformal vectors in vertex operator algebras", *J. Algebra* **179**:2 (1996), 523–548. MR 96m:17052 Zbl 0964.17021

[Miyamoto 2004] M. Miyamoto, "A new construction of the moonshine vertex operator algebra over the real number field", *Ann. of Math.* (2) **159**:2 (2004), 535–596. MR 2005h:17052 Zbl 1133.17017

[Nakanishi and Tsuchiya 1992] T. Nakanishi and A. Tsuchiya, "Level-rank duality of WZW models in conformal field theory", *Comm. Math. Phys.* **144**:2 (1992), 351–372. MR 93a:81181 Zbl 0751.17024

[Sakuma 2007] S. Sakuma, "6-transposition property of τ-involutions of vertex operator algebras", *Int. Math. Res. Not.* **2007**:9 (2007), Art. ID # rnm030. MR 2008h:17033 Zbl 1138.17013

[Wilson 1988] R. A. Wilson, "Some subgroups of the Thompson group", *J. Austral. Math. Soc. Ser. A* **44**:1 (1988), 17–32. MR 88k:20039 Zbl 0641.20016

Received May 13, 2013. Revised January 27, 2014.

HSIAN-YANG CHEN
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI 10617
TAIWAN
hychen@math.sinica.edu.tw

CHING HUNG LAM
INSTITUTE OF MATHEMATICS
ACADEMIA SINICA
TAIPEI 10617
TAIWAN
and
NATIONAL CENTER FOR THEORETICAL SCIENCES
NATIONAL CHENG KUNG UNIVERSITY
TAINAN 701
TAIWAN
chlam@math.sinica.edu.tw

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.53

SOFIC GROUPS: GRAPH PRODUCTS AND GRAPHS OF GROUPS

LAURA CIOBANU, DEREK F. HOLT AND SARAH REES

We prove that graph products of sofic groups are sofic, as are graphs of groups for which vertex groups are sofic and edge groups are amenable.

1. Introduction

We prove the following results.

Theorem 1.1. *A graph product of sofic groups is sofic.*

Theorem 1.2. *The fundamental group of a graph of groups is sofic if each vertex group is sofic and each edge group is amenable.*

Theorem 1.1 generalizes Theorem 1 of [Elek and Szabó 2006], and our proof is based on ideas used in the proof of that theorem. Theorem 1.2 is an extension of the result that free products of sofic groups amalgamated over amenable subgroups are sofic, proved independently in [Elek and Szabó 2011, Theorem 1] and [Păunescu 2011, Corollary 2.3]; most of the argument needed to extend the result is already found in [Collins and Dykema 2011, Corollary 3.6].

The term "sofic groups" is attributed to Weiss [2000] and applied to a definition due to Gromov [1999]; this is a class of groups which, together with the related class of hyperlinear groups, has inspired much recent study through its connections to a variety of different mathematical areas. A very useful introduction to sofic groups is provided by [Pestov 2008]. There are many open questions, including the question of whether all groups are sofic.

A number of quite distinct, but equivalent, definitions exist for sofic groups and are proved equivalent in [Pestov 2008]. The definition in [Weiss 2000] for finitely generated groups involves finite subsets of the Cayley graph of the group and is essentially the same as the definition in [Gromov 1999] of the Cayley graph being *initially subamenable*. An alternative and equivalent definition of [Pestov 2008] defines a group to be sofic if it embeds as a subgroup in an ultraproduct of symmetric groups. Another (equivalent) definition, found in [Elek and Szabó 2006], is phrased in terms of quasi-actions. We shall work with a variation of that

MSC2010: 20F65, 37B05.
Keywords: sofic, graph products, free and direct products, groups of graphs.

definition, given below as Definition 1.4; we phrase it in terms of (what we call) *special* quasi-actions. That this is equivalent to the definition of [Elek and Szabó 2006] (and hence to the others) follows from Lemma 2.1 of the same paper.

For a finite set A, let $\mathcal{S}(A)$ be the group of all permutations of A. For $\epsilon > 0$, we say that two elements f_1, f_2 of $\mathcal{S}(A)$ are ϵ-similar if the number of elements $a \in A$ for which $f_1(a) \neq f_2(a)$ is at most $\epsilon|A|$. Note that for $\epsilon \geq 1$ this condition is always satisfied.

Definition 1.3. Suppose that G is a group, $\epsilon > 0$ is a real number and $F \subseteq G$ is a finite subset of G. A special (F, ϵ)-quasi-action of G on a finite set A is a function $\phi : G \to \mathcal{S}(A)$ with the following properties:

(a) $\phi(1) = 1$.

(b) $\phi(g)^{-1} = \phi(g^{-1})$ for all $g \in G$.

(c) For $g \in F \setminus \{1\}$, $\phi(g)$ has no fixed points.

(d) For $g_1, g_2 \in F$ the map $\phi(g_1 g_2)$ is ϵ-similar to $\phi(g_1)\phi(g_2)$.

For $a \in A$, $g \in G$, we write $a^{\phi(g)}$ for the image of a under $\phi(g)$.

Definition 1.4. A group G is *sofic* if, for each number $\epsilon \in (0, 1)$ and any finite subset $F \subseteq G$, G admits a special (F, ϵ)-quasi-action.

It is immediate from the definition that a group is sofic precisely if every one of its finitely generated subgroups is sofic. We note at this stage also the following elementary result, which will be useful to us later.

Lemma 1.5. *Let ϕ_i be special (F, ϵ)-quasi-actions of G on A_i for $1 \leq i \leq n$, let $A = A_1 \times \cdots \times A_n$, and define $\phi : G \to \mathcal{S}(A)$ by $(a_1, \ldots, a_n)^{\phi(g)} = (a_1^{\phi_1(g)}, \ldots, a_n^{\phi_n(g)})$. Then ϕ is a special $(F, n\epsilon)$-quasi-action.*

Proof. The conditions (a), (b) and (c) of the definition are straightforward to check for ϕ. The equality $(a_1, \ldots, a_n)^{\phi(g_1)\phi(g_2)} = (a_1, \ldots, a_n)^{\phi(g_1 g_2)}$ holds whenever $a_i^{\phi_i(g_1)\phi_i(g_2)} = a_i^{\phi_i(g_1 g_2)}$ for each a_i, which is the case for at least $(1-\epsilon)^n|A|$ elements $(a_1, \ldots, a_n) \in A$. The result now follows since $(1-\epsilon)^n \geq 1 - n\epsilon$ for all $n \geq 1$. □

This article contains two additional sections; Section 2 contains the proof of Theorem 1.1 and Section 3 the proof of Theorem 1.2.

2. Proof of the graph product theorem

Let Γ be a simple graph and, for each vertex v of Γ, let G_v be a group. The graph product of the groups G_v with respect to Γ is defined to be the quotient of their free product by the normal closure of the relators $[g_v, g_w]$ for all $g_v \in G_v$, $g_w \in G_w$ for which $\{v, w\}$ is an edge of Γ. Graph products were introduced by Green [1990] in her Ph.D. thesis, and their basic properties are established there. For a graph product of vertex groups $G_1, \dots, G_n$ with respect to a finite graph Γ with vertices $1, \dots, n$, and for $J \subseteq \{1, \dots, n\}$, we define $G_J := \langle G_j \mid j \in J \rangle$. By [loc. cit., Proposition 3.31], G_J is isomorphic to the graph product of the groups G_j $(j \in J)$ on the full subgraph of Γ with vertex set J. Note that $G_\varnothing$ is the trivial group.

Green only considered graph products of finitely many vertex groups, but the definition applies equally well to graphs with infinite vertex sets I. Since any relation in a group is a consequence of finitely many defining relations, the property that, for any $J \subseteq I$, the group G_J is isomorphic to the graph product of G_j $(j \in J)$ on the full subgraph of Γ with vertex set J, extends to graph products with infinitely many vertex groups. Hence, since a group is sofic if and only if all of its finitely generated subgroups are sofic, it suffices to prove Theorem 1.1 for graph products of finitely many groups, so we shall assume from now on that the graph Γ is finite.

Any nonidentity element in a graph product can be written as a product $g_1 \cdots g_l$ for some $l > 1$, where each g_i is a nontrivial element of a vertex group G_{j_i}. By [Green 1990, Theorem 3.9], we can get from any such expression of minimal length to any other by swapping the order in the expression of elements g_i, g_{i+1} from commuting vertex groups. Hence every minimal length expression for an element g has the same length l, which we call the *syllable length* of g, and involves the same set $\{g_1, g_2, \dots, g_l\}$ of vertex group elements with the same multiplicities, the *syllables* of g. Whenever $g_1 \cdots g_l$ is a minimal length expression for g, we call each product $g_1 \cdots g_i$ a *left divisor* of g and each product $g_{i+1} \cdots g_n$ a *right divisor* of g for $0 \le i \le n$.

We also note that, for any finite subset of a graph product of groups G_i, there is a bound N on the syllable lengths of its elements, and there are finite subsets F_i of the vertex groups G_i that contain all the syllables of those elements. Hence Theorem 1.1 follows from the following proposition.

Proposition 2.1. *There is a function $f : \mathbb{N} \to \mathbb{N}$ with the following property. Let $G_1, \dots, G_n$ be sofic groups and G be their graph product with respect to a finite graph Γ. Let $\epsilon > 0$ be given, and for each $i = 1, \dots n$, let F_i be a finite subset of G_i, let A_i be a finite set, and suppose that $\psi_i : G_i \to \mathcal{S}(A_i)$ is a special (F_i, ϵ)-quasi-action of G_i on A_i.*

Then, for any $N \in \mathbb{N}$, G has a special $(F, f(n)\epsilon)$-quasi-action ϕ on a finite set C, where F is the set of elements of G of syllable length at most N for which each syllable is in some F_i, such that the following additional properties hold:

(1) *whenever x, y are in distinct vertex groups, $\phi(xy) = \phi(x)\phi(y)$;*

(2) *C admits equivalence relations $\sim_1, \dots, \sim_n$ such that, for each $c \in C$, $g \in F$ and $J \subseteq \{1, \dots, n\}$,*
$$c^{\phi(g)} \sim_J c \iff g \in G_J$$
(where $\sim_J$ is the join of those equivalence relations $\sim_j$ for which $j \in J$).

Note, by definition, $a \sim_J b$ if and only if there is a sequence $a = c_1, \ldots, c_m = b$ of elements with $c_i \sim_{j_i} c_{i+1}$ for some $j_i \in J$. In particular, $x \sim_\varnothing y \Longleftrightarrow x = y$.

Note that the conditions (1) and (2) imposed on the special quasi-action ϕ are necessary for the inductive proof of the proposition, rather than to deduce the theorem. Condition (1) ensures in particular that $\phi(x)\phi(y) = \phi(y)\phi(x)$ whenever x, y are from commuting vertex groups.

Proof. The proof is by induction on n. Suppose first that $n = 1$. Then $G = G_1$ and $F = F_1$ (for any value of $N \in \mathbb{N}$). We put $F := F_1$ and $C := A_1$, and define the equivalence relation $\sim_1$ by $c \sim_1 d$ for all $c, d \in C$. Then ϕ is a special (F, ϵ)-quasi-action on C, and the additional property (1) holds vacuously. To see that the additional property (2) also holds, note that there are only two possibilities for J: $J = \{1\}$ and $J = \varnothing$. If $J = \{1\}$, then $G = G_J$, so the left- and right-hand sides of the equivalence in (2) are true for all $g \in G$. If $J = \varnothing$ then, by the definition of a special (F, ϵ)-quasi-action, both the left and right hand sides of the equivalence are true if and only if $g = 1$. So the property (2) holds, and the statement of the proposition is true with $f(n) = 1$.

Now we proceed to prove the inductive step. We shall prove that the result holds with $f(n) = n(nf(n-1)+1)$.

Write $I = \{1, 2, \ldots, n\}$, and for each $k \in I$, let $I_k = I \setminus \{k\}$. For each $k \in I$, let $H_k := G_{I_k}$ be the subgroup of G that is the graph product of the groups G_i for $i \neq k$ with respect to the appropriate subgraph of Γ. By the induction hypothesis, we may assume that, for $\epsilon' := f(n-1)\epsilon$ and $F_{H_k} := F \cap H_k$, the subgroup H_k has a special (F_{H_k}, ϵ')-quasi-action θ_k on a set D_k admitting equivalence relations $\simeq_i^k$ for each $i \neq k$, such that

(1) $\theta_k(xy) = \theta_k(x)\theta_k(y)$ for x, y in distinct vertex groups of H_k; and

(2) for $d \in D_k$, $h \in F_{H_k}$, and $J \subseteq I_k$, we have $d^{\theta_k(h)} \simeq_J^k d \Longleftrightarrow h \in G_J$.

For each $k \in I$, we shall build a set C_k related to D_k, admitting equivalence relations $\sim_i^k$ for each $i \in I$, and then construct a special quasi-action ϕ_k of G on C_k that satisfies (1) and more. We shall then construct ϕ and the equivalence relations $\sim_1, \ldots, \sim_n$ on the set $C := C_1 \times C_2 \times \cdots \times C_n$ in terms of the special quasi-actions ϕ_k and the equivalence relations $\sim_i^k$ using Lemma 1.5.

For $k \in I$, let $L_k \subseteq I_k$ be the set of vertices joined in Γ to k. Let $\simeq_{L_k}$ be the join of the equivalence relations $\simeq_i^k$ for $i \in L_k$, and let π_k be the projection from D_k to its set of equivalence classes under $\simeq_{L_k}$ (for which the image of $d \in D_k$ is its equivalence class).

Now, using ideas from [Elek and Szabó 2006, Theorem 1], we choose a finite group V_k, with generating set $\pi_k(D_k) \times A_k$, for which all relators among the generators have length greater than N, and we let the $C_k := D_k \times A_k \times V_k$.

We define equivalence relations $\sim_i^k$ on C_k for $i \neq k$ by the rules

$$(d, a, v) \sim_i^k (d', a', v') \iff d \simeq_i^k d',\ a = a',\ v = v'.$$

Then we define $\sim_k^k$ on C_k by specifying its equivalence classes: for $d \in D_k$, $v \in V_k$, the class $\alpha_k(d, v)$ is the subset $\{(d, a, v \circ (\pi_k(d), a)) : a \in A_k\}$ of C_k. Multiplication $\circ$ within the third component is the group multiplication of V_k.

We define a special quasi-action ϕ_k of G on C_k as a composite of natural extensions to C_k of the special quasi-actions θ_k, ψ_k of H_k and G_k on D_k, A_k.

For $h \in H_k$, we define

$$(d, a, v)^{\phi_k(h)} = (d^{\theta_k(h)}, a, v).$$

Then, for $g \in G_k$, we define

$$(d, a, v)^{\phi_k(g)} = (d, a^{\psi_k(g)}, v \circ (\pi_k(d), a)^{-1} \circ (\pi_k(d), a^{\psi_k(g)})).$$

Now it follows, essentially from [Green 1990, Lemma 3.20], that each element $g \in G$ has a unique expression as a product $g = x_1 y_1 \cdots x_m y_m$, with each $x_i \in H_k$, each $y_i \in G_k$, x_i nontrivial for $i > 1$, y_i nontrivial for $i < m$, and such that, for $i > 1$, x_i has no nontrivial left divisor in the subgroup G_{L_k}; we call this expression the *normal form* for g (with respect to k). We note that the y_i are syllables, the x_i are products of syllables and the number of terms is at most the syllable length of g. We use that expression for g to extend to G the definitions of ϕ_k on H_k and G_k; that is, $\phi_k(g) := \phi_k(x_1)\phi_k(y_1) \cdots \phi_k(x_m)\phi_k(y_m)$ for $g \in G$.

We need now the following lemma, whose proof we defer.

Lemma 2.2. *Let $\epsilon'' := (nf(n-1)+1)\epsilon$. Then, for each k, the map ϕ_k is a special (F, ϵ'')-quasi-action of G on C_k, such that*

(1) *whenever x, y are in distinct vertex groups, $\phi_k(xy) = \phi_k(x)\phi_k(y)$;*

(2′) *for each $c \in C_k$, $g \in F$, we have $g \in G_J \Rightarrow c^{\phi_k(g)} \sim_J^k c$ for all $J \subseteq I$, and $c^{\phi_k(g)} \sim_J^k c \Rightarrow g \in G_J$ for all $J \subseteq I_k$.*

Let $C := C_1 \times \cdots \times C_n$. Now we define a map $\phi : G \to \mathcal{S}(C)$ by $(c_1, \ldots, c_n)^{\phi(g)} = (c_1^{\phi_1(g)}, \ldots, c_n^{\phi_n(g)})$. It follows from Lemma 1.5 that this is a $(F, f(n)\epsilon)$-quasi-action with $f(n) = n(nf(n-1)+1)$. Condition (1) of the proposition is inherited from the maps ϕ_k.

We define equivalence relations $\sim_1, \ldots, \sim_n$ on C by $(c_1, \ldots, c_n) \sim_j (c_1', \ldots, c_n')$ if and only if $c_k \sim_j^k c_k'$ for $1 \leq k \leq n$. We need now to verify condition (2).

Let $J \subseteq I$. The fact that $g \in G_J$ implies that $c^{\phi(g)} \sim_J c$ for all $c \in C$ is inherited from the maps ϕ_k. If $J = I$, then $G = G_J$ and the converse statement is immediate. Otherwise we have $J \subseteq I_k$ for some k with $1 \leq k \leq n$. If $g \notin G_J$ and $c = (c_1, \ldots, c_n) \in C$, then $c_k^{\phi_k(g)} \not\sim_J^k c_k$ and hence $c^{\phi(g)} \not\sim_J c$.

So the proof of the proposition will be complete once we prove Lemma 2.2. □

Proof of Lemma 2.2: Note that since θ_k is a special (F_{H_k}, ϵ')-quasi-action for H_k, it is clear that the restriction of ϕ_k to H_k is as well. Certainly that restriction preserves each of the $\sim_i^k$ equivalence classes with $i \neq k$. Since ψ_k is a special $(F \cap G_k, \epsilon)$-quasi-action for G_k, it is clear that the restriction of ϕ_k to G_k is as well. That restriction preserves the $\sim_k^k$ equivalence classes, since both (d, a, v) and $(d, a, v)^{\phi_k(g)}$ are in $\alpha_k(d, v \circ (\pi_k(d), a)^{-1})$.

The equation $(d, a, v)^{\phi_k(1_G)} = (d, a, v)$ follows immediately from $(d, a, v)^{\phi_k(h)} = (d^{\theta_k(h)}, a, v)$ for $h \in H_k$, and hence condition (a) of Definition 1.3 is verified for ϕ_k.

We shall verify the remaining conditions in the order (c), (1), (b), (d), (2′).

First we introduce some notation. We need to consider $\phi_k(g)$ for a general element g in the graph product, written in normal form as $x_1 y_1 \cdots x_m y_m$. For $0 \leq i \leq m$, we write $\Theta_k(x, i)$ for the product $\theta_k(x_1) \cdots \theta_k(x_i)$ and $\Psi_k(y, i)$ for the product $\psi_k(y_1) \cdots \psi_k(y_i)$, where $\Theta_k(x, 0) = \Psi_k(y, 0) = 1$.

We see then that

$$(d, a, v)^{\phi_k(g)} = (d, a, v)^{\phi_k(x_1)\phi_k(y_1)\cdots\phi_k(x_m)\phi_k(y_m)} = (d^{\Theta_k(x,m)}, a^{\Psi_k(y,m)}, v \circ u),$$

where

$$u = \prod_{i=1}^{m} (\pi_k(d^{\Theta_k(x,i)}), a^{\Psi_k(y,i-1)})^{-1} \circ (\pi_k(d^{\Theta_k(x,i)}), a^{\Psi_k(y,i)}),$$

unless y_m is the identity, in which case the product for u is from $i = 1$ to $m - 1$.

Our next step is to establish condition (c) of Definition 1.3 for ϕ_k. Let g be a nontrivial element of F with normal form $x_1 y_1 \cdots x_m y_m$. Then $2m \leq N$, and $x_i \in F_{H_k}$ and $y_i \in F_k$ for $1 \leq i \leq m$.

Suppose first that u, in the above expression, is not the empty word. Since ψ_k is a special quasi-action, condition (c) for ψ_k implies that $a^{\Psi_k(y,i-1)} \neq a^{\Psi_k(y,i)}$ for each i. Since $x_{i+1} \notin G_{L_k}$, it follows from the induction hypothesis that $\theta_k(x_{i+1})$ cannot map any element of D_k to an element in the same $\simeq_{L_k}$ equivalence class, and hence $\pi_k(d^{\Theta_k(x,i)}) \neq \pi_k(d^{\Theta_k(x,i+1)})$. Thus no generator in the word of length $2m$ representing u can freely cancel with the generator either before it or after it. The fact that V admits no short relators now ensures that u is nontrivial. In that case certainly $(d, a, v)^{\phi_k(g)} \neq (d, a, v)$.

Now suppose that u is empty. Then $m = 1$, y_1 is trivial and $g = x_1$. Since $1 \neq g$, we have $d^{\theta_k(x_1)} \neq d$ and again we have $(d, a, v)^{\phi_k(g)} = (d^{\theta_k(x_1)}, a, v) \neq (d, a, v)$.

Hence we have shown that the map ϕ_k from G to $\mathcal{S}(C_k)$ allows no nonidentity element of length less than N in F to fix any element of C_k, and so condition (c) is verified for ϕ_k.

In order to establish condition (1) of the lemma for ϕ_k, we suppose first that $x \in G_{L_k}$ and $y \in G_k$. By definition, $\phi_k(xy) = \phi_k(x)\phi_k(y)$, and

$$(d, a, v)^{\phi_k(x)\phi_k(y)} = (d^{\theta_k(x)}, a^{\psi_k(y)}, v \circ (\pi_k(d^{\theta_k(x)}), a)^{-1} \circ (\pi_k(d^{\theta_k(x)}), a^{\psi_k(y)})),$$

while

$$\begin{aligned}(d,a,v)^{\phi_k(y)\phi_k(x)} &= (d, a^{\psi_k(y)}, v\circ(\pi_k(d),a)^{-1})\circ(\pi_k(d),a^{\psi_k(y)}))^{\phi_k(x)} \\ &= (d^{\theta_k(x)}, a^{\psi_k(y)}, v\circ(\pi_k(d),a)^{-1}\circ(\pi_k(d),a^{\psi_k(y)})).\end{aligned}$$

Then since $d \simeq_{L_k} d^{\theta_k(x)}$, we have $\pi_k(d)=\pi_k(d^{\theta_k(x)})$, and so

$$(d,a,v)^{\phi_k(x)\phi_k(y)} = (d,a,v)^{\phi_k(y)\phi_k(x)};$$

that is, for $x \in G_{L_k}$, $y \in G_k$, we have $\phi_k(xy)=\phi_k(x)\phi_k(y)=\phi_k(y)\phi_k(x)$.

Now suppose that x, y are in distinct vertex groups G_i, G_j. If $i, j \neq k$ then condition (1) follows immediately by induction applied to G_k. If $j = k$, or if $i = k$ and G_i, G_j do not commute, then xy is in normal form, and condition (1) follows from the definition of ϕ_k. Finally if $i = k$ and G_i, G_j commute, then $x \in G_k$, $y \in G_{L_k}$, and we can deduce condition (1) for ϕ_k from the result above.

Next suppose that $g = x_1y_1\cdots x_my_m \in G$, where the expression is in normal form. We compare $\phi_k(g)^{-1}$ and $\phi_k(g^{-1})$. We have $g^{-1} = y_m^{-1}x_m^{-1}\cdots y_1^{-1}x_1^{-1}$. The expression for g^{-1} need not be in normal form because some of the x_i^{-1} could have left divisors in G_{L_k}, but we can put it into normal form by splitting any such x_i^{-1} into syllables and then applying commuting relations to move left divisors of x_i^{-1} in G_{L_k} past y_i^{-1}. By the results of the preceding two paragraphs, if we apply the corresponding transformations to $\phi_k(y_m^{-1})\phi_k(x_m^{-1})\cdots\phi_k(y_1^{-1})\phi_k(x_1^{-1})$, then we do not change the resulting permutation. Hence $\phi_k(g^{-1}) = \phi_k(y_m^{-1})\phi_k(x_m^{-1})\cdots\phi_k(y_1^{-1})\phi_k(x_1^{-1})$. It follows from condition (b) of Definition 1.3 that $\phi_k(y_i^{-1})$ is inverse to $\phi_k(y_i)$ and from the induction hypothesis on H_k that $\phi_k(x_i^{-1})$ is inverse to $\phi_k(x_i)$. Hence $\phi_k(g^{-1}) = \phi_k(g)^{-1}$, which verifies condition (b) for ϕ_k.

We proceed now to verify condition (d) of Definition 1.3 for ϕ_k, that is, to show that $\phi_k(g_1g_2)$ is ϵ''-similar to $\phi_k(g_1)\phi_k(g_2)$ for all $g_1, g_2 \in F$. Let $g_1 = x_1y_1\cdots x_my_m$, $g_2 = x'_1y'_1\cdots x'_py'_p$ be the normal forms of $g_1, g_2 \in F$. In the following discussion, we refer to an element of H_k or of G_k as a *block*, and to a product of blocks as an *expression*. The normal form for g_1g_2 is derived from the concatenation $x_1y_1\cdots x_my_mx'_1y'_1\cdots x'_py'_p$ by a sequence of moves, each of which is one of four types:

(a) deletion of a block that is equal to the identity;

(b) cancellation (that is, merger of two adjacent mutually inverse blocks that are either both in H_k or both in G_k);

(c) expression of a block in H_k as a product of a left divisor in G_{L_k} and a right divisor, and moving the left divisor to the left, past a block in G_k;

(d) merger of two adjacent blocks that are either both in H_k or both in G_k, and whose product is not the identity, to give a new block from that same subgroup.

Note that in (c) the left and right divisors of a block in H_k are simply subblocks whose concatenation is a permutation of the original block; that is, the (multi)set of syllables of the block in H_k is the union of the (multi)sets of syllables of those left and right divisors. By contrast, a move of type (d) will normally change the (multi)set of syllables in an expression. Starting with the permutation

$$\phi_k(x_1)\phi_k(y_1)\cdots\phi_k(x_m)\phi_k(y_m)\phi_k(x'_1)\phi_k(y'_1)\cdots\phi_k(x'_p)\phi_k(y'_p),$$

we study the sequence of composites of permutations of C_k defined by the various expressions that arise when we apply the corresponding operations to this expression of images during this rewrite process and keep track of the proportion of elements of C_k on which they differ. We note that, as a consequence of what we have proved so far, two expressions that differ only on moves of types (a), (b) and (c) correspond to composites of permutations that have the same effect on all points of C_k. Hence we only need to concern ourselves with moves of type (d).

Suppose that a move converts an expression w to an expression w'. Let σ, σ' be the permutations corresponding to the two expressions. If the move merges two blocks from G_k, then the permutations σ and σ' differ on the same proportion of elements of C_k as do permutations for the quasi-action of G_k on the set A_k, that is, on at most $\epsilon|C_k|$ of the elements, by the hypothesis.

If the move merges two blocks from H_k, then the permutations σ and σ' differ on the same proportion of elements of C_k as do permutations for the quasi-action of H_k on the set D_k, that is, on at most $f(n-1)\epsilon|C_k|$ of the elements by the induction hypothesis. Notice however that if the two blocks z_1, z_2 being merged are left and right divisors of z_1z_2 (or, equivalently, if the syllable length of z_1z_2 is the sum of the syllable lengths of z_1 and z_2), then our induction hypothesis on H_k ensures that $\phi_k(z_1z_2) = \phi_k(z_1)\phi_k(z_2)$. We shall call such mergers *nonreducing*, and other mergers, for which this equality is not guaranteed to hold, *reducing*. Condition (d) of Definition 1.3 can now be now established by applying the following lemma.

Lemma 2.3. *During the rewrite process, we perform at most n reducing mergers of blocks of H_k and at most one reducing merger of blocks of G_k.*

Proof. We may assume that $m, p > 0$ (since otherwise one of g_1, g_2 is the identity) and split the proof into three cases (1) $1 \neq y_m$ and $x'_1 \notin G_{L_k}$; (2) $y_m = 1$; and (3) $1 \neq y_m$ and $x'_1 \notin G_{L_k}$.

We deal with Case 1 first, proving by induction on m that in this case the product can be rewritten using at most $|L_k|$ mergers, all of which are within H_k. Using that result we deal with the remaining two cases together, also using induction on m.

Case 1: $1 \neq y_m$ *and* $x'_1 \notin G_{L_k}$. Let $x'_1 = z_1z_2$, where z_1 is the longest left divisor of x'_1 in G_{L_k}. Suppose that $z_1 \in G_{L'}$ for some $L' \subseteq L_k$. We prove by induction on m that this product can be rewritten using at most $|L'|$ $(\le |L_k|)$ H_k-mergers and no G_k-mergers.

If $m = 1$ then there can be at most one H_k-merger $x_1 z_1$, so the result is clear. So suppose that $m > 1$; then $y_{m-1} \neq 1$, and x_m is nontrivial with no left divisor in G_L. If z_1 commutes with x_m, then the claim follows by induction applied to the product $(x_1 y_1 \cdots x_{m-1} y_{m-1})(z_1 x_m y_m z_2 y'_1 \cdots x'_p y'_p)$. Otherwise, we can write $z_1 = z_{11} z_{12}$, where z_{11} (which may be trivial) is the longest left divisor of z_1 that commutes with x_m. So $z_{11} \in G_{L''}$ with $|L''| < |L'|$. We can then perform the rewriting by performing an H_k-merger $x_m z_{12}$ (if necessary) and, by induction, at most $|L''|$ further H_k-mergers resulting from moving z_{11} further to the left. This completes the proof of the claim and of the lemma in Case 1.

So now we may assume that $m, p > 0$, and that we are in Case 2 or 3.

<u>Case 2</u>: $y_m = 1$. If $m = 1$, then there is at most one H_k-merger $x_1 x'_1$, and the result holds. So suppose that $m > 1$ and hence that $y_{m-1} \neq 1$, and x_m is nontrivial with no left divisor in G_{L_k}.

If $x_m x'_1 \notin G_{L_k}$, then we perform an H_k-merger (if necessary) on $x_m x'_1$, and now observe that the product $(x_1 y_1 \cdots x_{m-1} y_{m-1})(x_m x'_1 y'_1 \cdots x'_p y'_p)$ satisfies the conditions of Case 1 and thus can be rewritten using at most $|L_k|$ further H_k-mergers and no G_k-mergers. So in this case too, the lemma is proved.

If $x_m x'_1 \in G_{L_k}$ then, since x_m has no left divisor in G_{L_k}, the product $x_m x'_1$ can be evaluated by writing x_m and x'_1 as products of syllables and then performing commuting and cancellation moves only so we can rewrite $x_m x'_1$ as $z \in G_{L_k}$, without performing any mergers, to arrive at the product

$$(x_1 y_1 \cdots x_{m-1} y_{m-1})(z y'_1 \cdots x'_p y'_p),$$

which satisfies the conditions of Case 3 for $m - 1$. The lemma now follows by induction applied to that product.

<u>Case 3</u>: $1 \neq y_m$ *and* $x'_1 \in G_{L_k}$. If $y_m y'_1 \neq 1$, we perform the G_k-merger $y_m y'_1$ and the H_k-merger $x_m x'_1$ (which cannot be in G_{L_k} since $x_m \notin G_{L_k}$, $x'_1 \in G_{L_k}$). Then we apply the result of Case 1 to the product $(x_1 y_1 \cdots x_{m-1} y_{m-1})(x_m x'_1 y_m y'_1 \cdots x'_p y'_p)$, and the proof is complete.

If $y_m y'_1 = 1$ then the result is clear if $p = 1$ and otherwise, since x'_2 has no left divisor in G_{L_k}, the merger $x'_1 x'_2$ is nonreducing, so the result follows by applying Case 2 to the product $(x_1 y_1 \cdots x_{m-1} y_{m-1} x_m)(x'_1 x'_2 y'_2 \cdots x'_p y'_p)$. □

This completes the proof of condition (d), and hence we see that ϕ_k is a special (F, ϵ'')-quasi-action, with $\epsilon'' = (nf(n-1)+1)\epsilon$.

It remains to verify condition (2′). We have shown already that, for each $i \in I$, the action of $\phi_k(G_i)$ on C preserves each of the $\sim_i^k$-equivalence classes, from which it follows immediately that $g \in G_J$ with $J \subseteq I$ implies $c^{\phi_k(g)} \sim_J^k c$.

Now suppose that $J \subseteq I_k$, $c = (d, a, v) \in C_k$, $g \in F$, and that $c^{\phi_k(g)} \sim_J^k c$. Since $k \notin J$, it is immediate from the definition of $\sim_j^k$ for $j \in J$ that

$$(d, a, v) \sim_J^k (d', a', v') \iff d \simeq_J^k d',\ a = a',\ v = v'.$$

Now, arguing as in our earlier proof of condition (c) of Definition 1.3 for ϕ_k that $(d, a, v)^{\phi_k(g)} \neq (d, a, v)$ for $1 \neq g \in F$, we find that, for $g \in F$, we have $(d, a, v)^{\phi_k(g)} \sim_J^k (d, a, v)$ if and only if $g \in H_k$ and $d^{\theta_k(g)} \simeq_J^k d$. By our inductive hypothesis, this is true if and only if $g \in G_J$. Hence condition (2′) holds. □

3. Graphs of groups

In this section we prove Theorem 1.2. We start by recalling the definition of a graph of groups, which arises from the work of Bass and Serre (see [Serre 1977]).

Definition 3.1. A graph of groups $\mathcal{G}$ consists of

(1) a connected graph Γ (in which loops are allowed, but no multiple edges), with vertex set V, edge set E,

(2) a collection of *vertex groups* $G_v : v \in V$ and *edge groups* $G_e : e \in E$,

(3) monomorphisms $\theta_e^1 : G_e \to G_{v_1}$ and $\theta_e^2 : G_e \to G_{v_2}$ for each edge $e = \{v_1, v_2\}$.

The *fundamental group* $\pi_1(\mathcal{G})$ of a graph of groups $\mathcal{G}$ can defined in various different (but equivalent) ways. The following definition is essentially [Dicks and Dunwoody 1989, Definition I.3.4]. The definition is given in terms of a selected spanning tree T of Γ, but (up to isomorphism) the resulting group is independent of this choice. The associated fundamental group $\pi_1(\mathcal{G}, T)$ is then the group generated by the groups $G_v : v \in V$ together with generators t_e, one for each (oriented) edge in E), given the following relations:

(1) all the relations of the groups G_v,

(2) $t_e^{-1}\theta_e^1(g)t_e = \theta_e^2(g)$ for each $e \in E$, $g \in G_e$,

(3) $t_e = 1$ for each edge e of T.

From this description it is not hard to see that $\pi_1(\mathcal{G}, T)$ is isomorphic to a multiple HNN extension, with stable letters t_e for $e \notin E(T)$, of the amalgamated product of the groups G_v in which $\theta_e^1(g)$ and $\theta_e^2(g)$ are identified for all $e \in E(T)$, $g \in G_e$. Independent results of Elek and Szabó [2011, Theorem 1] and Păunescu [2011, Corollary 2.3] already prove that the amalgamated product of two sofic groups over an amenable subgroup is sofic. Hence Theorem 1.2 follows immediately by combining that result with the following proposition.

Proposition 3.2. *An HNN extension of a sofic group H over an amenable subgroup K is sofic.*

We deduce Proposition 3.2 as a corollary of the amalgamated product result. We note that the argument to do this was already provided by Collins and Dykema [2011] in order to deduce Corollary 3.6 from their Theorem 3.4, that is to deduce the same result as above in the situation where the associated subgroups (in both amalgamated products and HNN extensions) are monotileably amenable. This argument goes through without any modification, when monotileability of the associated subgroup is dropped, to deduce the proposition from the results of [Elek and Szabó 2011; Păunescu 2011], but we include the argument here for completeness.

Proof. Let G be an HNN extension of H over K, as in the proposition, and let L be the subgroup $t^{-1}Kt$. Define $H_i = t^{-i}Ht^i$, $K_i = t^{-i}Kt^i$, $L_i = t^{-i}Lt^i$ for each $i \in \mathbb{Z}$, and define $S := \langle H_i \mid i \in \mathbb{Z} \rangle$. Then G can be expressed as an extension of S by $\mathbb{Z}$. Since $\mathbb{Z}$ is amenable, and by [Elek and Szabó 2006, Theorem 1(3)] an extension of a sofic group by an amenable group is sofic, in order to prove G is sofic it is enough to prove S is sofic.

Now S can be expressed as an iterated amalgamated product of the (countably many) H_i with amalgamation over subgroups isomorphic to K. More precisely, S is the fundamental group of the graph of groups $\mathcal{H}$ associated with the graph of the integers, where H_i is the vertex group of the vertex i, each edge group is isomorphic to K, and the copy of K associated with edge $\{i, i+1\}$ maps to the subgroup L_i of H_i and the subgroup K_{i+1} of H_{i+1}. Here is a diagram of $\mathcal{H}$:

$$\cdots H_{i-1} \xrightarrow{L_{i-1} \hookleftarrow \quad \hookrightarrow K_i} H_i \xrightarrow{L_i \hookleftarrow \quad \hookrightarrow K_{i+1}} H_{i+1} \xrightarrow{L_{i+1} \hookleftarrow \quad \hookrightarrow K_{i+2}} H_{i+2} \cdots$$

To prove S is sofic we now need to verify soficity for each of its finitely generated subgroups. So let M be such a subgroup. Then, for some k, l, all the generators of M are within vertex subgroups H_i for $k \le i \le l$; that is, M is a subgroup of the amalgamated product $H_j *_{L_j = K_{j+1}} H_{j+1} *_{L_{j+1} = K_{j+2}} * \cdots *_{L_{l-1} = K_l} H_l$. Since this is sofic, by [Elek and Szabó 2011; Păunescu 2011], so is M. □

Acknowledgments

The authors would like to thank the referee for their careful reading of the paper and for pointing out some significant technical errors in the original version. All three authors were partially supported by the Marie Curie Reintegration Grant 230889. Ciobanu was also supported by the Swiss National Science Foundation grants Ambizione PZ00P-136897/1 and Professorship FN PP00P2-144681/1.

References

[Collins and Dykema 2011] B. Collins and K. J. Dykema, "Free products of sofic groups with amalgamation over monotileably amenable groups", *Münster J. Math.* **4** (2011), 101–117. MR 2869256 Zbl 1242.43003

[Dicks and Dunwoody 1989] W. Dicks and M. J. Dunwoody, *Groups acting on graphs*, Cambridge Studies in Advanced Mathematics **17**, Cambridge University Press, 1989. MR 91b:20001 Zbl 0665.20001

[Elek and Szabó 2006] G. Elek and E. Szabó, "On sofic groups", *J. Group Theory* **9**:2 (2006), 161–171. MR 2007a:20037 Zbl 1153.20040

[Elek and Szabó 2011] G. Elek and E. Szabó, "Sofic representations of amenable groups", *Proc. Amer. Math. Soc.* **139**:12 (2011), 4285–4291. MR 2012j:20127 Zbl 1263.43001

[Green 1990] E. R. Green, *Graph products of groups*, thesis, University of Leeds, 1990, Available at http://etheses.whiterose.ac.uk/236.

[Gromov 1999] M. Gromov, "Endomorphisms of symbolic algebraic varieties", *J. Eur. Math. Soc. (JEMS)* **1**:2 (1999), 109–197. MR 2000f:14003 Zbl 0998.14001

[Păunescu 2011] L. Păunescu, "On sofic actions and equivalence relations", *J. Funct. Anal.* **261**:9 (2011), 2461–2485. MR 2012j:46089 Zbl 1271.46051 arXiv 1002.0605

[Pestov 2008] V. G. Pestov, "Hyperlinear and sofic groups: a brief guide", *Bull. Symbolic Logic* **14**:4 (2008), 449–480. MR 2009k:20103 Zbl 1206.20048

[Serre 1977] J.-P. Serre, *Arbres, amalgames,* SL_2, Astérisque **46**, Société Mathématique de France, Paris, 1977. Translated as *Trees*, Springer, Berlin, 2003. MR 57 #16426 Zbl 0369.20013

[Weiss 2000] B. Weiss, "Sofic groups and dynamical systems", *Sankhyā Ser. A* **62**:3 (2000), 350–359. MR 2001j:37022 Zbl 1148.37302

Received December 12, 2012. Revised November 14, 2013.

LAURA CIOBANU
MATHEMATICS DEPARTMENT
UNIVERSITY OF NEUCHÂTEL
RUE EMILE-ARGAND 11
CH-2000 NEUCHÂTEL
SWITZERLAND
laura.ciobanu@unine.ch

DEREK F. HOLT
MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY
CV4 7AL
UNITED KINGDOM
D.F.Holt@warwick.ac.uk

SARAH REES
SCHOOL OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEWCASTLE
NEWCASTLE
NE1 7RU
UNITED KINGDOM
Sarah.Rees@newcastle.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.65

PERTURBATIONS OF A CRITICAL FRACTIONAL EQUATION

EDUARDO COLORADO, ARTURO DE PABLO AND URKO SÁNCHEZ

We deal with the following fractional critical problem:

$$\begin{cases} (-\Delta)^{\alpha/2}u = |u|^{2\alpha/(N-\alpha)}u + f(x) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

where $\Omega \subset \mathbb{R}^N$ is a regular bounded domain, $0 < \alpha < 2$ and $N > \alpha$. Under appropriate conditions on the size of f, we prove existence and multiplicity of solutions.

1. Introduction

It is well known, using the Pohozhaev identity [1970], that the critical problem

$$\begin{cases} -\Delta u = |u|^{4/(N-2)}u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega \end{cases} \tag{1-1}$$

has no positive solution whenever Ω is a star-shaped domain. Starting from this nonexistence result, in the last decades several perturbations of this problem have been investigated in order to obtain a solution and understand the criticality of the problem. A pioneering work in that sense is the one performed by Brézis and Nirenberg [1983], in which the authors study the existence of positive solutions of the problem

$$\begin{cases} -\Delta u = |u|^{4/(N-2)}u + f(x,u) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1-2}$$

where $f(x,u) = f(u) = \lambda u$, with $\lambda \in \mathbb{R}$ and $N > 2$. Among other extensions, we highlight the work [Ambrosetti et al. 1994], where the authors studied the case $f(u) = \lambda|u|^{q-2}u$ with $1 < q < 2$, as well as [Tarantello 1992], in which the case $f(x,u) = f(x)$ was investigated; see also [Rey 1992].

Our purpose here is to study the similar situation that occurs for the fractional Laplacian in a bounded domain and the corresponding critical power.

E. Colorado is partially supported by Spanish Research Projects Ref. MTM2009-10878 and MTM2010-18128. A. de Pablo is partially supported by Spanish Research Project Ref. MTM2011-25287.
MSC2010: 35A15, 49J35, 35R11.
Keywords: semilinear elliptic equations, fractional Laplacian, critical problem.

We define the fractional Laplacian in a bounded domain Ω via its spectral decomposition, namely

$$(-\Delta)^{\alpha/2}u = \sum a_j \rho_j^{\alpha/2} \varphi_j,$$

where $\{\rho_j, \varphi_j\}$ is the spectral decomposition of the operator $-\Delta$ in Ω under zero Dirichlet boundary conditions and the a_j are the coefficients of u for the base $\{\varphi_j\}$ in $L^2(\Omega)$. A more precise notation would be $(-\Delta)_\Omega^{\alpha/2}$, because the operator strongly depends on the domain Ω, but we omit the subscript since the domain is fixed throughout the paper. We also recall that in the case where the domain under consideration is the whole space $\mathbb{R}^N$, the associated fractional Laplacian operator $(-\Delta)_{\mathcal{F}}^{\alpha/2}$ is defined via Fourier transformation for functions in the Schwartz class:

$$\big[(-\Delta)_{\mathcal{F}}^{\alpha/2} g\big]^\wedge(\xi) = |\xi|^\alpha \hat{g}(\xi),$$

which gives a different operator.

The critical problem corresponding to (1-2) with the fractional Laplacian is

$$\begin{cases} (-\Delta)^{\alpha/2}u = |u|^{2\alpha/(N-\alpha)}u + f(x,u) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{1-3}$$

First, there is again a (fractional) Pohozhaev-type identity, which in the case $f \equiv 0$ yields, as for the classical problem (1-1), the nonexistence of positive solutions whenever Ω is a star-shaped domain; see [Brändle et al. 2013]. Note that

$$\frac{2\alpha}{N-\alpha} = 2_\alpha^* - 2, \quad \text{where} \quad 2_\alpha^* := \frac{2N}{N-\alpha}$$

is the critical Sobolev exponent associated to α.

Next, in the case $f(x,u) = f(u)$, we point out [Barrios et al. 2012], in which an existence and multiplicity result was proved for positive solutions when

$$f(u) = \lambda |u|^{q-2}u, \quad \lambda > 0, \quad 0 < \alpha < 2, \quad N > \alpha, \quad \text{and} \quad 1 < q < \frac{2N}{N-\alpha}.$$

The case $\alpha = 1$ and $q = 2$ was studied previously in [Tan 2011].

In this paper we investigate zero order perturbations, $f(x,u) = f(x)$ small in (1-3), of the critical problem $f \equiv 0$, in relation to the results of [Tarantello 1992] for the classical Laplace operator. Thus, we consider the following problem:

$$\begin{cases} (-\Delta)^{\alpha/2}u = |u|^{p-2}u + f(x) & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{P}$$

where $0 < \alpha < 2$, $N > \alpha$, $p = \dfrac{2N}{N-\alpha}$ and f belongs to a suitable space.

In order to establish the functional setting, we define the function space $H_0^{\alpha/2}(\Omega)$ as the completion of $\mathscr{C}_0^\infty(\Omega)$ endowed with the norm

$$\|u\|_{H_0^{\alpha/2}} = \|(-\Delta)^{\alpha/4}u\|_2 = \Bigl(\sum a_j^2\rho_j^{\alpha/2}\Bigr)^{1/2}.$$

The operator $L(u) = (-\Delta)^{\alpha/2}u - |u|^{p-2}u$ is well defined from $H_0^{\alpha/2}(\Omega)$ into its dual $H^{-\alpha/2}(\Omega)$ by the Sobolev inequality; see (2-3) below. Thus it is natural to consider data f in that space: we have $f \in H^{-\alpha/2}(\Omega)$ if and only if $f = (-\Delta)^{\alpha/2}g$ with $g \in H_0^{\alpha/2}(\Omega)$; the associated norm is given by $\|f\|_{H^{-\alpha/2}} = \|g\|_{H_0^{\alpha/2}}$.

Throughout, we will consider solutions of the problem (P) in the following sense:

Definition 1.1. Let $f \in H^{-\alpha/2}(\Omega)$. We say that $u \in H_0^{\alpha/2}(\Omega)$ is an energy solution to the problem (P) if

$$\int_\Omega (-\Delta)^{\alpha/4}u(x)(-\Delta)^{\alpha/4}\psi(x)\,dx = \int_\Omega (|u(x)|^{p-2}u(x) + f(x))\psi(x)\,dx \tag{1-4}$$

for every $\psi \in H_0^{\alpha/2}(\Omega)$.

In the sequel we use the simplified notation $\int f = \int f(x)\,dx$ when no confusion can arise.

The paper is organized as follows: In Section 2 we state the main existence results and establish some preliminaries, and Sections 3 and 4 contain the proofs.

2. Main results and preliminaries

We will focus on functions $f \in H^{-\alpha/2}(\Omega)$ that are small in the following sense:

$$\int_\Omega f\varphi < c(\alpha, N)\|\varphi\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha} \quad \text{for all } \varphi \in H_0^{\alpha/2}(\Omega) \text{ with } \|\varphi\|_p = 1, \tag{2-1}$$

where

$$c(\alpha, N) = \frac{2\alpha}{N-\alpha}\Bigl(\frac{N-\alpha}{N+\alpha}\Bigr)^{\frac{N+\alpha}{2\alpha}}.$$

The main result of the paper is the following:

Theorem 2.1. *Assume $f \not\equiv 0$ satisfies* (2-1). *Then the problem* (P) *has at least two solutions. Moreover, if $f \ge 0$ a.e. in Ω, then these solutions are nonnegative a.e. in Ω.*

We will also prove that, if we relax the strict inequality in condition (2-1) by replacing it with the condition

$$\int_\Omega f\varphi \le c(\alpha, N)\|\varphi\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha} \quad \text{for all } \varphi \in H_0^{\alpha/2}(\Omega) \text{ with } \|\varphi\|_p = 1, \tag{2-2}$$

then we still obtain the existence of at least one solution:

Theorem 2.2. *Assume $f \not\equiv 0$ satisfies* (2-2)*. Then the problem* (P) *has at least one solution. Moreover, if f is nonnegative a.e. in Ω then this solution is nonnegative a.e. in Ω.*

For the fractional Laplacian defined above and $N > \alpha$, the following Sobolev inequality holds:

$$\text{(2-3)} \quad \int_\Omega |(-\Delta)^{\alpha/4}\varphi|^2 \geq S(\alpha, N)\left(\int_\Omega |\varphi|^{2N/(N-\alpha)}\right)^{\frac{N-\alpha}{N}} \quad \text{for all } \varphi \in H_0^{\alpha/2}(\Omega).$$

See, for example, [Brändle et al. 2013], where the inequality is proved as a consequence of the Hardy–Littlewood–Sobolev inequality [Hardy and Littlewood 1928; Sobolev 1938]. In the case of $\mathbb{R}^N$ and $(-\Delta)^{\alpha/4}_{\mathcal{F}}$, it takes the form

$$\text{(2-4)} \quad \int_{\mathbb{R}^N} |(-\Delta)^{\alpha/4}_{\mathcal{F}}\varphi|^2 \geq S(\alpha, N)\left(\int_{\mathbb{R}^N} |\varphi|^{2N/(N-\alpha)}\right)^{\frac{N-\alpha}{N}} \quad \text{for all } \varphi \in \mathcal{S}(\mathbb{R}^N).$$

The value of $S(\alpha, N)$ can be seen, for instance, in [Lieb 1983]. It is independent of the domain and is not attained in any bounded domain, although it is attained in $\mathbb{R}^N$.

The condition (2-1) is equivalent to

$$\text{(2-5)} \qquad \int_\Omega f\varphi < c(\alpha, N)\frac{\|\varphi\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha}}{\|\varphi\|_p^{N/\alpha}} \quad \text{for all } \varphi \in H_0^{\alpha/2}(\Omega)\backslash\{0\}.$$

Moreover, since

$$\text{(2-6)} \qquad \int_\Omega f\varphi \leq \|f\|_{H^{-\alpha/2}}\|\varphi\|_{H_0^{\alpha/2}},$$

using the Sobolev inequality (2-3) we obtain the following sufficient condition for f to satisfy (2-1):

$$\text{(2-7)} \qquad \|f\|_{H^{-\alpha/2}} \leq c(\alpha, N)S(\alpha, N)^{N/2\alpha}.$$

Remarks. (1) An assumption on the size of f like (2-1) is necessary in order to find solutions of problem (P). For example, if f is a sufficiently large positive constant, then problem (P) has no solutions.

(2) Condition (2-7) seems not to be sharp, in view of the result in [Castro and Zuluaga 1993] for the case $\alpha = 2$, which could also be proved in our functional framework.

The associated energy functional to problem (P) is given by

$$I(u) = \frac{1}{2}\int_\Omega \left|(-\Delta)^{\alpha/4}u\right|^2 - \frac{1}{p}\int_\Omega |u|^p - \int_\Omega fu.$$

Clearly, *critical points* of I correspond to solutions of (P) in the sense of (1-4). Indeed, one of the solutions we will construct in the proof of Theorem 2.1 is a local minimum of I in $H_0^{\alpha/2}(\Omega)$.

3. Proof of Theorem 2.1

First solution. We start with the definition of the Nehari manifold associated to problem (P):

$$\mathcal{N} = \{u \in H_0^{\alpha/2}(\Omega) : u \not\equiv 0,\ \langle I'(u), u\rangle = 0\}.$$

It is natural to look for solutions in this manifold. Note that the condition $u \in \mathcal{N}$ is equivalent to the identity

$$\|u\|^2_{H_0^{\alpha/2}} = \|u\|_p^p + \int_\Omega fu. \tag{3-1}$$

Therefore the functional I restricted to $\mathcal{N}$ takes the equivalent forms

$$I(u) = \frac{\alpha}{2N}\|u\|^2_{H_0^{\alpha/2}} - \frac{N+\alpha}{2N}\int_\Omega fu = \frac{\alpha}{2N}\|u\|_p^p - \frac{1}{2}\int_\Omega fu. \tag{3-2}$$

We will use both expressions in the sequel. In particular, using the first one we deduce that the functional I is bounded from below on $\mathcal{N}$:

$$I(u) \ge \frac{\alpha}{2N}\|u\|^2_{H_0^{\alpha/2}} - \frac{N+\alpha}{2N}\|f\|_{H^{-\alpha/2}}\|u\|_{H_0^{\alpha/2}} \ge -\frac{(N+\alpha)^2}{8N\alpha}\|f\|^2_{H^{-\alpha/2}}, \tag{3-3}$$

where the last step is a consequence of the minimization of the function

$$\alpha t^2 - (N+\alpha)\|f\|_{H^{-\alpha/2}}\, t.$$

Remark. Taking (3-3) into account, it makes sense to define

$$c_0 = \inf\nolimits_{\mathcal{N}} I > -\infty, \tag{3-4}$$

although the functional is not bounded from below in the whole space $H_0^{\alpha/2}(\Omega)$.

Note that if u_0 is a local minimum of I in $H_0^{\alpha/2}(\Omega)$, then necessarily

$$\|u_0\|^2_{H_0^{\alpha/2}} - (p-1)\|u_0\|_p^p \ge 0.$$

In fact, as we will prove in Lemma 3.4, this inequality is strict; namely,

$$\|u_0\|^2_{H_0^{\alpha/2}} - (p-1)\|u_0\|_p^p > 0. \tag{3-5}$$

In the same way, if u_0 is a local maximum of I, we have

$$\|u_0\|^2_{H_0^{\alpha/2}} - (p-1)\|u_0\|_p^p < 0. \tag{3-6}$$

Thus, we first minimize the restriction of the functional I to $\mathcal{N}$ in order to find a critical point and therefore a solution to the problem (P). As we will see, c_0 is achieved. To prove that, we start with some preliminary results.

Lemma 3.1. *Let $f \not\equiv 0$ satisfy* (2-1). *Given $u \in H_0^{\alpha/2}(\Omega)$, assume $\int_\Omega fu > 0$. Then there exist unique numbers $\sigma = \sigma(u) > 0$ and $\tau = \tau(u) > \sigma$ with $\sigma u,\ \tau u \in \mathcal{N}$ and such that* (3-5) *is satisfied with $u_0 = \sigma u$ and* (3-6) *with $u_0 = \tau u$.*

Proof. Let $\theta(t) = t\|u\|^2_{H_0^{\alpha/2}} - t^{p-1}\|u\|_p^p$. The maximum value of this function occurs at

$$t_M = \left(\frac{(N-\alpha)\|u\|^2_{H_0^{\alpha/2}}}{(N+\alpha)\|u\|_p^p}\right)^{\frac{N-\alpha}{2\alpha}},$$

and

$$\theta(t_M) = \frac{2\alpha}{N-\alpha}\left(\frac{N-\alpha}{N+\alpha}\right)^{\frac{N+\alpha}{2\alpha}} \frac{\|u\|^{(N+\alpha)/\alpha}_{H_0^{\alpha/2}}}{\|u\|_p^{N/\alpha}} = c(\alpha, N)\frac{\|u\|^{(N+\alpha)/\alpha}_{H_0^{\alpha/2}}}{\|u\|_p^{N/\alpha}}.$$

Note that θ is a concave function, increasing on $(0, t_M)$ and decreasing on (t_M, ∞), with $\lim_{t\to\infty}\theta(t) = -\infty$. By (2-5) we get $0 < \int_\Omega fu < \theta(t_M)$. Thus there exist two unique values $0 < \sigma < t_M < \tau$ such that

$$\theta(\tau) = \int_\Omega fu = \theta(\sigma), \quad \theta'(\tau) < 0 < \theta'(\sigma). \tag{3-7}$$

Multiplying in the previous expression by τ, we have

$$0 = \tau\theta(\tau) - \tau\int_\Omega fu = \|\tau u\|^2_{H_0^{\alpha/2}} - \|\tau u\|_p^p - \int_\Omega \tau f u;$$

thus $\tau u \in \mathcal{N}$. Moreover,

$$\|\tau u\|^2_{H_0^{\alpha/2}} - (p-1)\|\tau u\|_p^p = \tau^2\theta'(\tau) < 0.$$

Arguing in a similar way for σ, we obtain $\sigma u \in \mathcal{N}$ and

$$\|\sigma u\|^2_{H_0^{\alpha/2}} - (p-1)\|\sigma u\|_p^p = \sigma^2\theta'(\sigma) > 0. \qquad \square$$

Observe that without the condition $\int_\Omega fu > 0$ we still can find a value $\tau > 0$ with $\tau u \in \mathcal{N}$ satisfying (3-5). Conversely, the condition $\int_\Omega fu > 0$ is guaranteed for any function $u \in \mathcal{N}$ that satisfies (3-5).

We notice that the purpose of the strict condition (2-1) on f is just to obtain $\int_\Omega fu < \theta(t_M)$. It also appears to be of importance in Lemma 3.3 below. It is known that, when one deals with the problem associated to the standard Laplacian, and under certain hypotheses, the condition (2-1) is not sharp; see [Castro and Zuluaga 1993]. We suspect that a similar fact can occur in our case.

Corollary 3.2. *Under the hypotheses of Lemma 3.1, we have*

$$I(\tau u) = \max_{t \geq \sigma} I(tu) \quad \textit{and} \quad I(\sigma u) = \min_{0 \leq t \leq \tau} I(tu).$$

Proof. It is straightforward once we notice that the function $g(t) = I(tu)$ satisfies $g'(t) = \theta(t) - \int_\Omega fu$. □

The next property uses a technical result analogous to Lemma 2.2 in [Tarantello 1992]. The proof follows almost word by word the one in that paper; see also [Brézis and Nirenberg 1989]. We only have to adapt the calculations to the functional framework of the fractional Laplacian. We leave the details to the interested reader.

Lemma 3.3. *Let $f \not\equiv 0$ satisfy* (2-1). *Then*

$$\mu_0 := \inf_{\substack{u \in H_0^{\alpha/2}(\Omega) \\ \|u\|_p = 1}} \left(c(\alpha, N) \|u\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha} - \int_\Omega fu \right) \tag{3-8}$$

is achieved, and moreover $\mu_0 > 0$.

The proof of this lemma is a straightforward adaptation to our setting of the similar one in the classical case; see Lemma 2.2 in [Tarantello 1992], which is inspired by the corresponding result in [Brézis and Nirenberg 1989].

The following lemma establishes a crucial property for minima of the functional; see inequality (3-5).

Lemma 3.4. *Let $f \not\equiv 0$ satisfy* (2-1) *and let $u \in \mathcal{N}$. Then*

$$\|u\|_{H_0^{\alpha/2}}^2 - (p-1)\|u\|_p^p \neq 0.$$

Proof. Consider the functional, defined for $u \in H_0^{\alpha/2}(\Omega)$, $u \not\equiv 0$, by

$$\phi(u) = c(\alpha, N) \frac{\|u\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha}}{\|u\|_p^{N/\alpha}} - \int_\Omega fu.$$

If $\|u\|_p = 1$, we have

$$\phi(tu) = t\left(c(\alpha, N) \|u\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha} - \int_\Omega fu \right).$$

Thus, given $\gamma > 0$ (to be chosen later), by Lemma 3.3 we have

$$\inf_{\|u\|_p \geq \gamma} \phi(u) \geq \gamma \mu_0. \tag{3-9}$$

Note that this infimum is also positive.

Now we suppose, for a contradiction, that there exists $u \in \mathcal{N}$ such that

$$\|u\|_{H_0^{\alpha/2}}^2 - (p-1)\|u\|_p^p = 0. \tag{3-10}$$

By the Sobolev inequality (2-3), we obtain

$$S(\alpha, N)\|u\|_p^2 - (p-1)\|u\|_p^p \le 0,$$

which implies

$$\|u\|_p \ge \Big(\frac{S(\alpha, N)}{p-1}\Big)^{1/(p-2)} =: \gamma.$$

Now, substituting (3-10) into (3-1), we get

$$0 = \|u\|^2_{H_0^{\alpha/2}} - \|u\|_p^p - \int_\Omega fu = (p-2)\|u\|_p^p - \int_\Omega fu\,. \tag{3-11}$$

Finally, by (3-9) and (3-11), we conclude that

$$\begin{aligned} 0 < \gamma\mu_0 \le \phi(u) &= (p-2)\Big(\frac{N-\alpha}{N+\alpha}\Big)^{\frac{N+\alpha}{2\alpha}} \frac{\|u\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha}}{\|u\|_p^{N/\alpha}} - \int_\Omega fu \\ &= (p-2)\Bigg[\Big(\frac{N-\alpha}{N+\alpha}\Big)^{\frac{N+\alpha}{2\alpha}} \frac{\|u\|_{H_0^{\alpha/2}}^{(N+\alpha)/\alpha}}{\|u\|_p^{N/\alpha}} - \|u\|_p^p\Bigg] \\ &= (p-2)\|u\|_p^p\Bigg[\bigg(\frac{(N-\alpha)\|u\|^2_{H_0^{\alpha/2}}}{(N+\alpha)\|u\|_p^p}\bigg)^{\frac{N-\alpha}{2\alpha}} - 1\Bigg] = 0, \end{aligned}$$

which is a contradiction. □

Lemma 3.5. *Let $f \not\equiv 0$ be a function satisfying* (2-1)*. Given $u \in \mathcal{N}$, there exists a positive function $\mu_u : H_0^{\alpha/2}(\Omega) \to \mathbb{R}$, differentiable in a neighborhood $\mathcal{U}_0$ of the origin in $H_0^{\alpha/2}(\Omega)$, such that*

$$\mu_u(0) = 1, \quad \mu_u(z)(u-z) \in \mathcal{N},$$

and

$$\langle \mu_u'(0), z\rangle = \frac{2\int_\Omega(-\Delta)^{\alpha/4}u(-\Delta)^{\alpha/4}z - p\int_\Omega |u|^{p-2}uz - \int_\Omega fz}{\|u\|^2_{H_0^{\alpha/2}} - (p-1)\|u\|_p^p} \tag{3-12}$$

for all $z \in \mathcal{U}_0$.

Proof. Consider the function

$$F(\mu, z) = \mu\|u-z\|^2_{H_0^{\alpha/2}(\Omega)} - \mu^{p-1}\|u-z\|_p^p - \int_\Omega f(u-z).$$

By Lemma 3.4, we have

$$\frac{\partial F}{\partial\mu}(1,0) = \|u\|^2_{H_0^{\alpha/2}} - (p-1)\|u\|_p^p \neq 0.$$

We complete the proof by applying the implicit function theorem to the function F at the point $(1, 0)$. □

We are now in a position to prove one of the main results of the paper.

Proposition 3.6. *The functional I possesses a local minimum in $H_0^{\alpha/2}(\Omega)$. In particular*, (P) *has a solution. Moreover*, *if f is nonnegative a.e. in Ω*, *this solution is nonnegative a.e. in Ω.*

Proof. Consider v, the unique solution to the equation $(-\Delta)^{\alpha/2}v = f$ in $H_0^{\alpha/2}(\Omega)$. Let $\sigma = \sigma(v)$ be as defined in Lemma 3.1. Since $\sigma(v)v \in \mathcal{N}$, we have

$$I(\sigma v) = \frac{\sigma^2}{2}\|v\|^2_{H_0^{\alpha/2}} - \frac{\sigma^p}{p}\|v\|_p^p - \sigma\|v\|^2_{H_0^{\alpha/2}} \tag{3-13}$$

$$= -\frac{\sigma^2}{2}\|v\|^2_{H_0^{\alpha/2}} + \frac{N+\alpha}{2N}\sigma^p\|v\|_p^p$$

$$< -\frac{\alpha\sigma^2}{2N}\|v\|^2_{H_0^{\alpha/2}} = -\frac{\alpha\sigma^2}{2N}\|f\|^2_{H^{-\alpha/2}}.$$

Then, by (3-3) and (3-13), the infimum in (3-4) satisfies the estimate

$$-\frac{(N+\alpha)^2}{8N\alpha}\|f\|^2_{H^{-\alpha/2}} \le c_0 < -\frac{\alpha\sigma^2}{2N}\|f\|^2_{H^{-\alpha/2}} < 0. \tag{3-14}$$

The expression (3-2) shows that the restriction of the functional I to $\mathcal{N}$ is weakly lower semicontinuous. Therefore, by Ekeland's variational principle [1974], we obtain a minimizing sequence of the functional I constrained to $\mathcal{N}$, i.e., $\{u_n\} \subset \mathcal{N}$ such that, for every $n \in \mathbb{N}$,

$$\text{(i)}\ \ I(u_n) < c_0 + \frac{1}{n} \quad \text{and} \quad \text{(ii)}\ \ \frac{1}{n}\|u_n - v\|_{H_0^{\alpha/2}} \ge I(u_n) - I(v) \ \text{ for all } v \in \mathcal{N}.$$

Combining (i), (3-14) and (3-2), we have

$$I(u_n) = \frac{\alpha}{2N}\|u_n\|^2_{H_0^{\alpha/2}} - \frac{N+\alpha}{2N}\int_\Omega f u_n < c_0 + \frac{1}{n} < -\frac{\alpha\sigma^2}{2N}\|f\|^2_{H^{-\alpha/2}}$$

for n large enough. Therefore

$$\frac{\alpha\sigma^2}{N+\alpha}\|f\|^2_{H^{-\alpha/2}} \le \int_\Omega f u_n \quad \text{and} \quad \|u_n\|^2_{H_0^{\alpha/2}} \le \frac{N+\alpha}{\alpha}\int_\Omega f u_n. \tag{3-15}$$

These inequalities, together with (2-6), give

$$\frac{\alpha\sigma^2}{N+\alpha}\|f\|_{H^{-\alpha/2}} \le \|u_n\|_{H_0^{\alpha/2}} \le \frac{N+\alpha}{\alpha}\|f\|_{H^{-\alpha/2}}. \tag{3-16}$$

Thus, we have (for a subsequence) that $u_n \rightharpoonup u_0$ weakly in $H^{\alpha/2}(\Omega)$, with $u_0 \not\equiv 0$. We claim that $\|I'(u_0)\|_{H^{-\alpha/2}} = 0$. Take $z \in H_0^{\alpha/2}(\Omega)$ with $\|z\|_{H_0^{\alpha/2}} = 1$. By Lemma 3.5, for every $n \in \mathbb{N}$, there exists a positive function μ_{u_n} such that

$$w_\delta = \mu_{u_n}(\delta z)(u_n - \delta z) \in \mathcal{N}$$

for $\delta > 0$ small enough. Set $t_n(\delta) = \mu_{u_n}(\delta z)$. Putting $v = w_\delta$ in (ii) and using the mean value theorem, we have

$$\frac{1}{n}\|w_\delta - u_n\|_{H_0^{\alpha/2}} \geq (1-t_n(\delta))\langle I'(w_\delta), u_n\rangle + \delta t_n(\delta)\langle I'(w_\delta), z\rangle + o(\delta).$$

Dividing by δ and taking the limit as δ goes to 0, we have

$$\frac{1}{n}\big(1+|t_n'(0)|\,\|u_n\|_{H_0^{\alpha/2}}\big) \geq \|I'(u_n)\|_{H^{-\alpha/2}}$$

with $|t_n'(0)| = \langle \mu_{u_n}'(0), z\rangle$, so that, by (3-16), we get

$$\|I'(u_n)\|_{H^{-\alpha/2}} \leq \frac{1}{n}\Big(1+\frac{N+\alpha}{\alpha}|t_n'(0)|\,\|f\|_{H^{-\alpha/2}}\Big). \tag{3-17}$$

Thus we are done once we prove that $|t_n'(0)|$ is uniformly bounded. By Lemma 3.5 and (3-16) we obtain

$$|t_n'(0)| \leq \frac{C}{\big|\|u_n\|^2_{H_0^{\alpha/2}} - (p-1)\|u_n\|_p^p\big|}$$

for some constant C. Assume, for a contradiction, that

$$\|u_n\|^2_{H_0^{\alpha/2}} - (p-1)\|u_n\|_p^p \to 0 \quad \text{as } n\to\infty. \tag{3-18}$$

From (3-18) and (3-1) we deduce the estimate

$$\int_\Omega f u_n = (p-2)\|u_n\|_p^p + o(1).$$

Moreover, from (3-16) we derive that $\|u_n\|_p \geq \gamma$ for some constant $\gamma > 0$. Thus, reasoning as in Lemma 3.4, we get

$$\begin{aligned} 0 < \gamma^{(N+\alpha)/2}\mu_0 &\leq \|u_n\|^{\alpha/N}_{H_0^{\alpha/2}}\phi(u_n) \\ &= (p-2)\Bigg[\bigg(\frac{(N-\alpha)\|u_n\|^2_{H_0^{\alpha/2}}}{N+\alpha}\bigg)^{\frac{N-\alpha}{2\alpha}} - (\|u_n\|_p^p)^{(N-\alpha)/2\alpha}\Bigg] \to 0, \end{aligned}$$

which leads to a contradiction. Therefore $\|I'(u_0)\|_{H^{-\alpha/2}} = 0$, and we have obtained a weak solution of (P).

To obtain strong convergence, we proceed as usual. Recalling that I is weakly lower semicontinuous in $\mathcal{N}$, we get

$$c_0 \leq I(u_0) \leq \lim_{n\to\infty} I(u_n) = c_0.$$

This implies, using (3-2), the limits

$$\lim_{n\to\infty}\|u_n\|_{H_0^{\alpha/2}} = \|u_0\|_{H_0^{\alpha/2}}, \qquad \lim_{n\to\infty}\|u_n\|_p = \|u_0\|_p.$$

To see that u_0 is a local minimum in $H_0^{\alpha/2}(\Omega)$, we first show that (3-5) holds. In fact, since $u_0 \in \mathcal{N}$ and also $\int_\Omega f u_0 > 0$ by (3-15), it is clear that one of the values $\sigma(u_0)$

or $\tau(u_0)$ given by Lemma 3.1 equals 1. Assume, for a contradiction (see Lemma 3.4), that u_0 satisfies (3-6), i.e., $\sigma(u_0) < \tau(u_0) = 1$. By Corollary 3.2, $I(\sigma(u_0)u_0) < I(u_0)$, which contradicts the fact that u_0 is the infimum in $\mathcal{N}$. Hence u_0 satisfies (3-5) and $\sigma(u_0) = 1$. We remark that having the strict inequality in (2-5) is crucial in the present argument. In particular, we have obtained $1 = \sigma(u_0) < t_M < \tau(u_0)$, or

$$1 < \left(\frac{(N-\alpha)\|u_0\|^2_{H_0^{\alpha/2}}}{(N+\alpha)\|u_0\|_p^p} \right)^{\frac{N-\alpha}{2\alpha}}, \tag{3-19}$$

which is the same. Take $\varepsilon > 0$ small enough such that

$$1 < \left(\frac{(N-\alpha)\|u_0 - z\|^2_{H_0^{\alpha/2}}}{(N+\alpha)\|u_0 - z\|_p^p} \right)^{\frac{N-\alpha}{2\alpha}} =: t_{M,\varepsilon} \tag{3-20}$$

for $\|z\|_{H_0^{\alpha/2}} < \varepsilon$. By Lemma 3.5, there exists a positive function $\mu_{u_0} : H_0^{\alpha/2}(\Omega) \to \mathbb{R}$ such that $\mu_{u_0}(z)(u_0 - z) \in \mathcal{N}$ for every $\|z\|_{H_0^{\alpha/2}} < \varepsilon$, with ε smaller if necessary. Indeed, by continuity we have $\mu_{u_0}(z) < t_{M,\varepsilon}$ for $\varepsilon > 0$ sufficiently small. Thus we get that $\mu_{u_0}(z)(u_0 - z)$ satisfies (3-5), and as a consequence of Lemma 3.1 and Corollary 3.2 applied to $u_0 - z$, we obtain

$$I(s(u_0 - z)) \geq I(\mu_{u_0}(z)(u_0 - z)) \geq I(u_0) \quad \text{for all } s \in (0, t_{M,\varepsilon}).$$

Since by (3-20) we can take $s = 1$, we conclude that $I(u_0 - z) \geq I(u_0)$ for every $\|z\|_{H_0^{\alpha/2}} < \varepsilon$, i.e, u_0 is a local minimum in $H_0^{\alpha/2}(\Omega)$.

To finish, we assume that $f \geq 0$. Then it follows that $\int_\Omega f|u_0| > 0$. Take $\sigma = \sigma(|u_0|) > 0$ and $\tau = \tau(|u_0|) > \sigma$. We have

$$\|u_0\|_p^p + \int_\Omega f u_0 = \|u_0\|^2_{H_0^{\alpha/2}} > (p-1)\|u_0\|_p^p$$

and, since $\tau|u_0|$ satisfies (3-6), we get

$$\tau^p \|u_0\|_p^p + \tau \int_\Omega f|u_0| = \tau^2 \big\| |u_0| \big\|^2_{H_0^{\alpha/2}} < (p-1)\tau^p \|u_0\|_p^p.$$

Thus,

$$(p-2)\|u_0\|_p^p < \int_\Omega f u_0 \leq \int_\Omega f|u_0| \leq (p-2)\tau^{p-1}\|u_0\|_p^p,$$

which implies $\tau > 1$. Therefore, by Corollary 3.2, we have

$$I(u_0) \leq I(\sigma|u_0|) \leq I(|u_0|).$$

On the other hand, by the generalized Stroock–Varopoulos inequality [de Pablo et al. 2012], we have

$$\int_\Omega \big|(-\Delta)^{\alpha/4}|u_0|\big|^2 \leq \int_\Omega |(-\Delta)^{\alpha/4} u_0|^2,$$

which implies $I(|u_0|) \le I(u_0)$. As a consequence, $I(u_0) = I(|u_0|)$, $\sigma = 1$, and thus $|u_0| \in \mathcal{N}$ is a solution. $\square$

Second solution. We will look for the second solution using a classical approach that relies on the well-known mountain pass theorem; see [Ambrosetti and Rabinowitz 1973]. Recall that $\{u_n\} \subset H_0^{\alpha/2}(\Omega)$ is a Palais–Smale (PS for short) sequence of level c for I if $I(u_n) \to c$ and $\|I'(u_n)\|_{H^{-\alpha/2}} \to 0$ as $n \to \infty$. Moreover, we say that I satisfies a PS *condition of level c* (PS_c for short) if every PS sequence of level c for I has a convergent subsequence in $H_0^{\alpha/2}(\Omega)$. As is usual in critical problems, the functional I does not satisfy a global PS condition, i.e., a PS_c condition for every c. Our aim is to prove that I satisfies a PS_c condition for c below a precise critical level c^*. We define

$$c^* = c_0 + \frac{\alpha}{2N} S(\alpha, N)^{N/\alpha}. \tag{3-21}$$

This value, which is obtained in the next lemma, also appears in several other contexts, for instance when one applies the concentration-compactness principle to critical problems; see [Ambrosetti et al. 1994; Brézis and Nirenberg 1983; Hardy and Littlewood 1928; Lions 1985] for the standard case, and, for example, [Barrios et al. 2012] for the fractional case, and [Barrios et al. 2014; Servadei and Valdinoci ≥ 2014] for different nonlocal operators which include a different fractional Laplacian.

Lemma 3.7. *The functional I satisfies a local PS_c condition for any $c < c^*$.*

Proof. Let $\{u_n\} \subset H_0^{\alpha/2}(\Omega)$ be a PS sequence of level $c < c^*$. It is easy to check that the u_n are uniformly bounded in $H^{\alpha/2}(\Omega)$. Thus, there exists a subsequence (still denoted by u_n) such that $u_n \rightharpoonup z_0$ weakly in $H_0^{\alpha/2}(\Omega)$. As a consequence, $z_0 \in H_0^{\alpha/2}(\Omega)$ is a solution of (P).

We rewrite u_n as $u_n = u_0 + \phi_n$ with $\phi_n \to 0$. Applying the Brézis–Lieb lemma [1983] we get

$$\|u_n\|_p^p = \|u_0\|_p^p + \|\phi_n\|_p^p + o(1). \tag{3-22}$$

On one hand, by (3-22) and taking n large enough we have

$$\begin{aligned} c^* > I(u_n) &= I(u_0) + \frac{1}{2}\|\phi_n\|^2_{H_0^{\alpha/2}} - \frac{1}{p}\|\phi_n\|_p^p + o(1) \\ &\ge c_0 + \frac{1}{2}\|\phi_n\|^2_{H_0^{\alpha/2}} - \frac{1}{p}\|\phi_n\|_p^p + o(1). \end{aligned}$$

Hence, by the definition of c^* in (3-21), we obtain

$$\frac{1}{2}\|\phi_n\|^2_{H_0^{\alpha/2}} - \frac{1}{p}\|\phi_n\|_p^p < \frac{\alpha}{2N} S(\alpha, N)^{N/\alpha} + o(1). \tag{3-23}$$

Taking into account that $\{u_n\}$ is a PS sequence, we have

$$\begin{aligned}
o(1) = \langle I'(u_n), u_n\rangle &= \|u_n\|^2_{H_0^{\alpha/2}} - \|u_n\|_p^p - \int_\Omega f u_n \\
&= \|u_0\|^2_{H_0^{\alpha/2}} - \|u_0\|_p^p - \int_\Omega f u_0 + \|\phi_n\|^2_{H_0^{\alpha/2}} - \|\phi_n\|_p^p + o(1) \\
&= \langle I'(u_0), u_0\rangle + \|\phi_n\|^2_{H_0^{\alpha/2}} - \|\phi_n\|_p^p + o(1) \\
&= \|\phi_n\|^2_{H_0^{\alpha/2}} - \|\phi_n\|_p^p + o(1).
\end{aligned} \tag{3-24}$$

Now we want to prove that $\{\phi_n\}$ has a subsequence strongly converging to 0 in $H_0^{\alpha/2}(\Omega)$. Suppose, on the contrary, that there are $C, k > 0$ such that $\|\phi_n\|_{H_0^{\alpha/2}} \geq C$ for all $n \geq k$. Using (2-3) in (3-24), we get $\|\phi_n\|_p^{p-2} \geq S(\alpha, N) + o(1)$ and hence

$$\|\phi_n\|_p^p \geq S(\alpha, N)^{N/\alpha} + o(1). \tag{3-25}$$

From (3-23) and (3-25), we have

$$\begin{aligned}
\frac{\alpha}{2N} S(\alpha, N)^{N/\alpha} \leq \frac{\alpha}{2N}\|\phi_n\|_p^p + o(1) &= \frac{1}{2}\|\phi_n\|^2_{H_0^{\alpha/2}} - \frac{1}{p}\|\phi_n\|_p^p + o(1) \\
&< \frac{\alpha}{2N} S(\alpha, N)^{N/\alpha},
\end{aligned}$$

which is a contradiction. □

It is known (see, for instance, [Chen et al. 2006]) that the minimizers for the Sobolev inequality (2-4) are given by the two-parameter family of functions

$$u_{\varepsilon,x_0}(x) = \frac{\varepsilon^{(N-\alpha)/2}}{(|x-x_0|^2+\varepsilon^2)^{(N-\alpha)/2}}, \tag{3-26}$$

where $x_0 \in \mathbb{R}^N$, $\varepsilon > 0$. In what follows we will use the notation

$$A = \|u_{\varepsilon,x_0}\|_p, \quad B = \|(-\Delta)^{\alpha/4}_{\mathscr{F}} u_{\varepsilon,x_0}\|_2 = \left(\int_{\mathbb{R}^N} |\xi|^\alpha |\hat{u}_{\varepsilon,x_0}(\xi)|^2\, d\xi\right)^{1/2}. \tag{3-27}$$

Note that the last quantity defines a norm in the homogeneous fractional Sobolev space $\dot{H}^{\alpha/2}(\mathbb{R}^N)$. Both numbers A and B are clearly independent of ε and x_0, and moreover $B^2 = S(\alpha, N)A^2$.

Without loss of generality we may assume that $0 \in \Omega$. We define a cut-off function $\theta \in \mathscr{C}^\infty(\mathbb{R}^N)$ by $\theta(x) = \theta_0(|x|/\rho)$ with $\rho > 0$, where $\theta_0 \in \mathscr{C}^\infty(\mathbb{R})$ is a nonincreasing function satisfying

$$\theta_0(s) = 1 \quad \text{if } s \leq \tfrac{1}{2}, \quad \theta_0(s) = 0 \quad \text{if } s \geq 1.$$

We now recall that, if u_0 is the solution constructed in the previous subsection, we can find a set $\Sigma \subset \Omega$ of positive Lebesgue measure such that $u_0 \geq \nu > 0$ a.e.

in Σ (replace u_0 with $-u_0$ and f with $-f$ if necessary). For $x_0 \in \Sigma$, we set $\tilde{u}_{\varepsilon,x_0} = \theta u_{\varepsilon,x_0} \in H_0^{\alpha/2}(\Omega)$.

Proposition 3.8. *In the above notation, for a.e. $x_0 \in \Sigma$ there exists $\varepsilon^* = \varepsilon^*(x_0) > 0$ sufficiently small such that*

$$\sup_{t\geq 0} I(u_0 + t\tilde{u}_{\varepsilon,x_0}) < c^* \quad \textit{for all } 0 < \varepsilon < \varepsilon^*. \tag{3-28}$$

We observe that when one evaluates the functional in (3-28), one needs to evaluate $\|\tilde{u}_{\varepsilon,x_0}\|_{H_0^{\alpha/2}}$; i.e., one needs to evaluate the fractional Laplacian of a product of functions. This requires the use of a different, but equivalent, norm which does not involve directly the fractional Laplacian. It uses the so-called *α-harmonic extension* of [Caffarelli and Silvestre 2007] for $(-\Delta)_{\mathcal{F}}^{\alpha/4}$, adapted to the bounded domain setting in [Brändle et al. 2013; Cabré and Tan 2010; Stinga and Torrea 2010].

Consider the semi-infinite cylinder $\mathscr{C}_\Omega = \{(x, y) : x \in \Omega,\ y > 0\} \subset \mathbb{R}_+^{N+1}$ and its lateral boundary $\partial_L \mathscr{C}_\Omega = \partial\Omega \times (0, \infty)$. For a function $u \in H_0^{\alpha/2}(\Omega)$, we denote its α-harmonic extension to $\mathscr{C}_\Omega$ by $w = \mathrm{E}_\alpha(u)$, defined as the solution to the problem

$$\begin{cases} \operatorname{div}(y^{1-\alpha}\nabla w) = 0 & \text{in } \mathscr{C}_\Omega, \\ w = 0 & \text{on } \partial_L \mathscr{C}_\Omega, \\ w = u & \text{on } \Omega \times \{y = 0\}. \end{cases} \tag{3-29}$$

Then the equation

$$-\kappa_\alpha \lim_{y\searrow 0} \frac{\partial w}{\partial y} = (-\Delta)^{\alpha/2} u$$

holds, with κ_α a positive constant. Let $X_0^\alpha(\mathscr{C}_\Omega)$ be the completion of $\mathscr{C}_0^\infty(\Omega\times[0, \infty))$ under the norm

$$\|\phi\|_{X_0^\alpha} = \left(\kappa_\alpha \int_{\mathscr{C}_\Omega} y^{1-\alpha} |\nabla\phi(x, y)|^2 \, dx\, dy\right)^{1/2}.$$

In the case $\Omega = \mathbb{R}^N$ the corresponding space for functions in the upper half-space $\mathbb{R}_+^{N+1}$ is $X^\alpha(\mathbb{R}_+^{N+1})$, which can be defined in the same way with the integral extended to $\mathbb{R}_+^{N+1}$. The extension operator can be characterized also as a minimization of the X_0^α-norm (equivalently, the X^α norm) for all the functions with common trace at $y = 0$. Note that the extension operator is an isometry from $H_0^{\alpha/2}(\Omega)$ to $X_0^\alpha(\mathscr{C}_\Omega)$ and from $H^{\alpha/2}(\mathbb{R}^N)$ to $X^\alpha(\mathbb{R}_+^{N+1})$; that is,

$$\|\mathrm{E}_\alpha(\psi)\|_{X_0^\alpha} = \|\psi\|_{H_0^{\alpha/2}} \quad \text{for all } \psi \in H_0^{\alpha/2}(\Omega), \tag{3-30}$$

$$\|\mathrm{E}_\alpha(\psi)\|_{X^\alpha} = \|\psi\|_{\dot{H}^{\alpha/2}} \quad \text{for all } \psi \in \dot{H}^{\alpha/2}(\mathbb{R}^N). \tag{3-31}$$

This means that

$$\|w\|_{X_0^\alpha} \geq \|w(\,\cdot\,, 0)\|_{H_0^{\alpha/2}} \quad \text{for all } w \in X_0^\alpha(\mathscr{C}_\Omega) \tag{3-32}$$

and

$$\|w\|_{X^\alpha} \geq \|w(\cdot\,,0)\|_{H^{\alpha/2}} \quad \text{for all } w \in X^\alpha(\mathbb{R}^N). \tag{3-33}$$

See [Brändle et al. 2013] for the details.

We define the family $w_{\varepsilon,x_0} = E_\alpha(u_{\varepsilon,x_0})$, with u_{ε,x_0} given in (3-26). We want to find a family of modified minimizers in the extended space, by using a cutoff function in $\mathscr{C}_\Omega$. To do that we take

$$\phi(x,y) = \theta_0\left(\frac{(|x-x_0|^2+y^2)^{1/2}}{\rho}\right),$$

where θ_0 is defined above. With this notation we define $\tilde{w}_{\varepsilon,\,x_0} = \phi w_{\varepsilon,x_0} \in X_0^\alpha(\mathscr{C}_\Omega)$ and $\tilde{w}_{\varepsilon,\,x_0}(\cdot\,,0) = \tilde{u}_{\varepsilon,\,x_0}(\cdot)$.

In [Barrios et al. 2012, Lemma 3.8] the following estimates for $\tilde{w}_{\varepsilon,x_0}$ are proved:

$$\|\tilde{w}_{\varepsilon,x_0}\|^2_{X_0^\alpha} = \|w_{\varepsilon,x_0}\|^2_{X^\alpha} + O(\varepsilon^{N-\alpha}). \tag{3-34}$$

In view of (3-30), (3-32) and (3-34), we have

$$\|\tilde{u}_{\varepsilon,x_0}\|^2_{H_0^{\alpha/2}} \leq B^2 + O(\varepsilon^{N-\alpha}). \tag{3-35}$$

Moreover,

$$\|\tilde{u}_{\varepsilon,x_0}\|^p_p \geq A^p + O(\varepsilon^N). \tag{3-36}$$

We now state a result that will be useful in the proof of Proposition 3.8. Its proof follows the same arguments as in [Brézis and Nirenberg 1989], with obvious changes for our setting, so we omit the details.

Lemma 3.9. *Assume that $a, b > 0$ and that u_0, $\tilde{u}_{\varepsilon,x_0}$ are defined as above. For $t \in [a,b]$, we have*

$$\begin{aligned}\|u_0+t\tilde{u}_{\varepsilon,x_0}\|^p_p = \|u_0\|^p_p + t^p\|\tilde{u}_{\varepsilon,x_0}\|^p_p + pt\int_\Omega |u_0|^{p-2}u_0\tilde{u}_{\varepsilon,x_0} \\ + pt^{p-1}\int_\Omega |\tilde{u}_{\varepsilon,x_0}|^{p-2}\tilde{u}_{\varepsilon,x_0}u_0 + o(\varepsilon^{(N-\alpha)/2}).\end{aligned} \tag{3-37}$$

Proof of Proposition 3.8. On the one hand, since $I(u_0+t\tilde{u}_{\varepsilon,x_0})|_{t=0} = c_0 < c^*$, by a continuity argument we can find $t_0, \varepsilon_0 > 0$ both small enough such that

$$I(u_0+t\tilde{u}_{\varepsilon,x_0}) < c^* \quad \text{for all } t \in (0,t_0) \text{ and all } \varepsilon \in (0,\varepsilon_0).$$

On the other hand, by Lemma 3.9, together with (3-36) and the fact that A and B are independent of ε, we have

$$I(u_0+t\tilde{u}_{\varepsilon,x_0}) \to -\infty \text{ as } t \to \infty \quad \text{for all } \varepsilon > 0.$$

Hence there exists $t_1 > 0$ large enough that

$$I(u_0 + t\tilde{u}_{\varepsilon,x_0}) < c_0 < c^* \quad \text{for all } t \ge t_1 \text{ and all } \varepsilon \in (0, \varepsilon_0).$$

Thus, we just need to prove that there exists $\varepsilon^* \in (0, \varepsilon_0)$ such that

$$\sup_{t_0 \le t \le t_1} I(u_0 + t\tilde{u}_{\varepsilon,x_0}) < c^*$$

for every $0 < \varepsilon < \varepsilon^*$.

Take $t \in [t_0, t_1]$. Clearly, we have

$$I(u_0+t\tilde{u}_{\varepsilon,x_0}) = \frac{1}{2}\|u_0\|^2_{H_0^{\alpha/2}} + t\int_\Omega (-\Delta)^{\alpha/4} u_0 (-\Delta)^{\alpha/4}\tilde{u}_{\varepsilon,x_0} + \frac{t^2}{2}\|\tilde{u}_{\varepsilon,x_0}\|^2_{H_0^{\alpha/2}} - \frac{1}{p}\|u_0 + t\tilde{u}_{\varepsilon,x_0}\|_p^p - \int_\Omega f u_0 - t\int_\Omega f\tilde{u}_{\varepsilon,x_0}. \tag{3-38}$$

Since $S(\alpha, N)$ is attained for the function u_{ε,x_0}, substituting (3-35), (3-36) and (3-37) in (3-38) we have

$$\begin{aligned} I(u_0 + t\tilde{u}_{\varepsilon,x_0}) \le \frac{1}{2}\|u_0\|^2_{H_0^{\alpha/2}} + t\int_\Omega (-\Delta)^{\alpha/4} u_0 (-\Delta)^{\alpha/4}\tilde{u}_{\varepsilon,x_0} + \frac{t^2}{2}B^2 \\ - \frac{1}{p}\|u_0\|_p^p - \frac{t^p}{p}A^p - t\int_\Omega |u_0|^{p-2}u_0\tilde{u}_{\varepsilon,x_0} - t^{p-1}\int_\Omega |\tilde{u}_{\varepsilon,x_0}|^{p-1}u_0 \\ - \int_\Omega f u_0 - t\int_\Omega f\tilde{u}_{\varepsilon,x_0} + o(\varepsilon^{(N-\alpha)/2}). \end{aligned}$$

On the other hand, since u_0 is solution of (P), we get

$$I(u_0+t\tilde{u}_{\varepsilon,\, x_0}) \le I(u_0) + \frac{t^2}{2}B^2 - t^{p-1}\int_\Omega |\tilde{u}_{\varepsilon,\, x_0}|^{p-1}u_0 - \frac{t^p}{p}A^p + o(\varepsilon^{(N-\alpha)/2}). \tag{3-39}$$

Extending u_0 by zero outside Ω, we get

$$\begin{aligned} \int_\Omega |\tilde{u}_{\varepsilon,x_0}|^{p-1}u_0 &= \int_{\mathbb{R}^N} u_0(x)\theta^{p-1}(x)\frac{\varepsilon^{(N+\alpha)/2}}{(|x-x_0|^2+\varepsilon^2)^{(N+\alpha)/2}} \\ &= \varepsilon^{(N-\alpha)/2}\int_{\mathbb{R}^N} u_0(x)\theta^{p-1}(x)\frac{1}{\varepsilon^N}\eta\Big(\frac{x-x_0}{\varepsilon}\Big), \end{aligned}$$

with $\eta(x) = (|x|^2+1)^{-(N+\alpha)/2}$. Thus, there exists a constant $\nu > 0$ such that

$$\int_{\mathbb{R}^N} u_0(x)\theta^{p-1}(x)\frac{1}{\varepsilon^N}\eta\Big(\frac{x-x_0}{\varepsilon}\Big) \ge K\nu$$

for every $\varepsilon > 0$ sufficiently small, $x_0 \in \Sigma$ and $K = \int_{\mathbb{R}^N}\eta(x) < \infty$. Therefore

$$\int_\Omega |\tilde{u}_{\varepsilon,x_0}|^{p-1}u_0 = \varepsilon^{(N-\alpha)/2}K\nu + o(\varepsilon^{(N-\alpha)/2}). \tag{3-40}$$

Substituting (3-40) in (3-39), we have

$$I(u_0 + t\tilde{u}_{\varepsilon,x_0}) \le c_0 + \frac{t^2}{2}B^2 - t^{p-1}\varepsilon^{(N-\alpha)/2}K\nu - \frac{t^p}{p}A^p + o(\varepsilon^{(N-\alpha)/2}). \tag{3-41}$$

Let us now define the function

$$g(s)=\frac{s^2}{2}B^2-s^{p-1}\varepsilon^{(N-\alpha)/2}K\nu-\frac{s^p}{p}A^p \quad \text{for } s>0,$$

and let $s_\varepsilon>0$ be the point of global maximum, i.e.,

$$0=g'(s_\varepsilon)=s_\varepsilon B^2-(p-1)s_\varepsilon^{p-2}\varepsilon^{(N-\alpha)/2}K\nu-s_\varepsilon^{p-1}A^p. \tag{3-42}$$

We denote $S_0=(B^2/A^p)^{1/(p-2)}$. Note that $0<s_\varepsilon<S_0$ and $s_\varepsilon\to S_0$ as $\varepsilon\searrow 0$. Let $\delta_\varepsilon>0$ be such that $s_\varepsilon=S_0(1-\delta_\varepsilon)$. Since $B^2/A^p=S_0^{p-2}$, by (3-42) we have

$$\Big(\frac{B^{2(p-1)}}{A^p}\Big)^{1/(p-2)}(1-\delta_\varepsilon-(1-\delta_\varepsilon)^{p-1})-(p-1)S_0^{p-2}(1-\delta_\varepsilon)^{p-2}\varepsilon^{(N-\alpha)/2}K\nu=0,$$

which implies

$$(p-2)\Big(\frac{B^{2(p-1)}}{A^p}\Big)^{1/(p-2)}\delta_\varepsilon=(p-1)S_0^{p-2}\varepsilon^{(N-\alpha)/2}K\nu+o(\varepsilon^{(N-\alpha)/2}). \tag{3-43}$$

By (3-41) with $t=s_\varepsilon$ and (3-43), we have

$$\begin{aligned}I(u_0+s_\varepsilon\tilde u_{\varepsilon,x_0})&\le c_0+\frac{s_\varepsilon^2}{2}B^2-s_\varepsilon^{p-1}\varepsilon^{(N-\alpha)/2}K\nu-\frac{s_\varepsilon^p}{p}A^p+o(\varepsilon^{(N-\alpha)/2})\\&=c_0+\frac{S_0^2}{2}B^2-S_0^{p-1}\varepsilon^{(N-\alpha)/2}K\nu-\frac{S_0^p}{p}A^p+o(\varepsilon^{(N-\alpha)/2})\\&=c_0+\frac{\alpha}{2N}S(\alpha,N)^{N/\alpha}-S_0^{p-1}\varepsilon^{(N-\alpha)/2}K\nu+o(\varepsilon^{(N-\alpha)/2})\\&=c^*-S_0^{p-1}\varepsilon^{(N-\alpha)/2}K\nu+o(\varepsilon^{(N-\alpha)/2}).\end{aligned}$$

Taking ε sufficiently small, this finishes the proof. □

Lemma 3.10. *Assume $f\not\equiv 0$ satisfies* (2-1). *Then the functional I possesses a critical point different from u_0. In particular,* (P) *has a second solution. Moreover, if $f\ge 0$ a.e. in Ω then this solution is nonnegative a.e. in Ω.*

Proof. Set $\eta_{\varepsilon,M}=u_0+M\tilde u_{\varepsilon,x_0}$, with $0<\varepsilon<\varepsilon^*$ and $x_0\in\Sigma$ so that (3-28) holds. Assume that $M>0$ is large enough such that $I(\eta_{\varepsilon,M})<c_0$.

Now we set

$$\Gamma=\{\gamma:[0,1]\to H_0^{\alpha/2}(\Omega)\text{ such that }\gamma(0)=u_0,\ \gamma(1)=\eta_{\varepsilon,M}\}.$$

By Proposition 3.8 we have

$$c_0<c_1=\inf_{\gamma\in\Gamma}\max_{t\in[0,1]}I(\gamma(t))<c^*.$$

Thus, using the mountain pass theorem, we obtain a PS sequence of level c_1, and as a consequence of Lemma 3.7 we can find a critical point u_1 in $H_0^{\alpha/2}(\Omega)$ with energy level $c_1>c_0$, i.e., u_1 is a solution of (P) with $u_1\not\equiv u_0$.

To prove that the solution is positive in the case that $f \geq 0$, we set

$$\tilde{\mathcal{N}} := \{u \in \mathcal{N} : u \text{ satisfies (3-6)}\}$$

and $c_2 = \inf_{\tilde{\mathcal{N}}} I$. Is easy to see that, taking a larger M if necessary, we can assume

$$c_0 < c_2 \leq c_1 < c^*. \tag{3-44}$$

Now, using Ekeland's variational principle and following the steps of the proof of Proposition 3.6, we can obtain a PS sequence of level c_2. Again, Lemma 3.7 implies the existence of a solution $u_2 \in \mathcal{N}$ such that $I(u_2) = c_2$. Put $\tau = \tau(|u_2|) > 0$. Then $\tau|u_2| \in \tilde{\mathcal{N}}$. Finally, by Corollary 3.2,

$$\inf_{\tilde{\mathcal{N}}} I = I(u_2) = \max_{t \geq t_M} I(tu_2) \geq I(\tau u_2) \geq I(\tau|u_2|),$$

which finishes the proof. □

Remark. Note that u_2 could coincide with u_1.

4. Proof of Theorem 2.2

When f satisfies condition (2-2) instead of (2-1), we use an approximation argument.

Proof of Theorem 2.2. Consider a sequence of numbers $\{\varepsilon_k\}_{k\in\mathbb{N}} \subset (0, 1)$ such that $\varepsilon_k \searrow 0$ as $k \to \infty$, and define $f_k = (1 - \varepsilon_k) f$. Clearly f_k satisfies condition (2-1) for every $k \in \mathbb{N}$. We define I_k and $\mathcal{N}_k$ in a natural way:

$$I_k(u) = \frac{1}{2} \int_\Omega |(-\Delta)^{\alpha/4} u|^2 - \frac{1}{p} \int_\Omega |u|^p - \int_\Omega f_k u,$$
$$\mathcal{N}_k = \{u \in H_0^{\alpha/2}(\Omega) : u \not\equiv 0,\ \langle I_k'(u), u\rangle = 0\}.$$

Let $u_k \in \mathcal{N}_k$ be the local minimum found via Theorem 2.1, namely,

$$I_k(u_k) = \inf_{\mathcal{N}_k} I_k := c_k.$$

In particular, we have

$$\langle I_k'(u_k), z\rangle = 0 \quad \text{for all } z \in H_0^{\alpha/2}(\Omega), \tag{4-1}$$

and moreover

$$\|u_k\|^2_{H_0^{\alpha/2}} - \|u_k\|_p^p - \int_\Omega f_k u_k = 0, \tag{4-2}$$

which, by (2-3) and (2-6), implies that $\|u_k\|^2_{H_0^{\alpha/2}} < C$ for any $k \in \mathbb{N}$ and some constant $C > 0$ independent of k. Take $u \in \mathcal{N}$ satisfying (3-5). Then

$$\int_\Omega f_k u > 0 \quad \text{for all } k \in \mathbb{N}.$$

Applying Lemma 3.1 with $f = f_k$ and $\mathcal{N} = \mathcal{N}_k$, we find the values $0 < \sigma_k < t_{M_k} < \tau_k$ such that $\sigma_k u, \tau_k u \in \mathcal{N}_k$. Since u satisfies the inequality (3-5), we have $\tau_k > 1$. Thus, by Corollary 3.2 we have $I_k(\sigma_k u) \le I_k(u)$, which leads to

$$c_k \le I_k(\sigma_k u) \le I_k(u) \le I(u) + \varepsilon_k \|f\|_{H^{-\alpha/2}} \|u\|_{H_0^{\alpha/2}} \le I(u) + C\varepsilon_k.$$

In particular, $c_k \le c_0 + C\varepsilon_k$. Finally, reasoning as in (3-13) with $f = f_k$, we obtain

$$-\frac{(N+\alpha)^2}{8N\alpha}\|f\|^2_{H^{-\alpha/2}} < -\frac{(N+\alpha)^2}{8N\alpha}\|f_k\|^2_{H^{-\alpha/2}} \le c_k \le c_0 + C\varepsilon_k.$$

After passing to a subsequence, we can assume that c_k converges to some value c' such that

$$-\frac{(N+\alpha)^2}{8N\alpha}\|f\|^2_{H^{-\alpha/2}} \le c' \le c_0.$$

Moreover, since $\|u_k\|^2_{H_0^{\alpha/2}}$ is uniformly bounded, again for a subsequence if necessary, we have $u_k \rightharpoonup u^*$ weakly in $H_0^{\alpha/2}(\Omega)$. Then by (4-1) we have

$$\langle I'(u^*), z\rangle = 0 \quad \text{for all } z \in H_0^{\alpha/2}(\Omega),$$

and $I(u^*) \le c_0$. This implies $u^* \in \mathcal{N}$ and $I(u^*) = c_0$, which finishes the proof. The positivity of the solution when the datum f is taken nonnegative follows from the same argument as in the proof of Theorem 2.1. □

We finally remark that the solution constructed in this way is not necessarily a minimum of the functional. Therefore we cannot prove the mountain pass geometry in order to find a second solution.

References

[Ambrosetti and Rabinowitz 1973] A. Ambrosetti and P. H. Rabinowitz, "Dual variational methods in critical point theory and applications", *J. Funct. Anal.* **14** (1973), 349–381. MR 51 #6412 Zbl 0273.49063

[Ambrosetti et al. 1994] A. Ambrosetti, H. Brézis, and G. Cerami, "Combined effects of concave and convex nonlinearities in some elliptic problems", *J. Funct. Anal.* **122**:2 (1994), 519–543. MR 95g:35059 Zbl 0805.35028

[Barrios et al. 2012] B. Barrios, E. Colorado, A. de Pablo, and U. Sánchez, "On some critical problems for the fractional Laplacian operator", *J. Differential Equations* **252**:11 (2012), 6133–6162. MR 2911424 Zbl 1245.35034

[Barrios et al. 2014] B. Barrios, E. Colorado, R. Servadei, and F. Soria, "A critical fractional equation with concave-convex power nonlinearities", *Ann. Inst. H. Poincaré Anal. Non Linéaire* (2014). To appear.

[Brändle et al. 2013] C. Brändle, E. Colorado, A. de Pablo, and U. Sánchez, "A concave-convex elliptic problem involving the fractional Laplacian", *Proc. Roy. Soc. Edinburgh Sect. A* **143**:1 (2013), 39–71. MR 3023003 Zbl 06238478

[Brézis and Lieb 1983] H. Brézis and E. Lieb, "A relation between pointwise convergence of functions and convergence of functionals", *Proc. Amer. Math. Soc.* **88**:3 (1983), 486–490. MR 84e:28003 Zbl 0526.46037

[Brézis and Nirenberg 1983] H. Brézis and L. Nirenberg, "Positive solutions of nonlinear elliptic equations involving critical Sobolev exponents", *Comm. Pure Appl. Math.* **36**:4 (1983), 437–477. MR 84h:35059 Zbl 0541.35029

[Brézis and Nirenberg 1989] H. Brézis and L. Nirenberg, "A minimization problem with critical exponent and nonzero data", pp. 129–140 in *Symmetry in nature* (Pisa, 1989), edited by G. Bernardini, Scuola Normale Superiore, Pisa, 1989. Reprinted by Springer, Berlin, 2007. Zbl 0763.46023

[Cabré and Tan 2010] X. Cabré and J. Tan, "Positive solutions of nonlinear problems involving the square root of the Laplacian", *Adv. Math.* **224**:5 (2010), 2052–2093. MR 2011c:35106 Zbl 1198.35286

[Caffarelli and Silvestre 2007] L. Caffarelli and L. Silvestre, "An extension problem related to the fractional Laplacian", *Comm. Partial Differential Equations* **32**:7-9 (2007), 1245–1260. MR 2009k:35096 Zbl 1143.26002

[Castro and Zuluaga 1993] R. Castro and M. Zuluaga, "Existence results for a class of nonhomogeneous elliptic equations with critical Sobolev exponent", *Note Mat.* **13**:2 (1993), 269–276. MR 96b:35060 Zbl 0831.35060

[Chen et al. 2006] W. Chen, C. Li, and B. Ou, "Classification of solutions for an integral equation", *Comm. Pure Appl. Math.* **59**:3 (2006), 330–343. MR 2006m:45007a Zbl 1093.45001

[Ekeland 1974] I. Ekeland, "On the variational principle", *J. Math. Anal. Appl.* **47** (1974), 324–353. MR 49 #11344 Zbl 0286.49015

[Hardy and Littlewood 1928] G. H. Hardy and J. E. Littlewood, "Some properties of fractional integrals, I", *Math. Z.* **27**:1 (1928), 565–606. MR 1544927 JFM 54.0275.05

[Lieb 1983] E. H. Lieb, "Sharp constants in the Hardy–Littlewood–Sobolev and related inequalities", *Ann. of Math.* (2) **118**:2 (1983), 349–374. MR 86i:42010 Zbl 0527.42011

[Lions 1985] P.-L. Lions, "The concentration-compactness principle in the calculus of variations: the limit case, II", *Rev. Mat. Iberoamericana* **1**:2 (1985), 45–121. MR 87j:49012 Zbl 0704.49006

[de Pablo et al. 2012] A. de Pablo, F. Quirós, A. Rodríguez, and J. L. Vázquez, "A general fractional porous medium equation", *Comm. Pure Appl. Math.* **65**:9 (2012), 1242–1284. MR 2954615 Zbl 1248.35220

[Pohožaev 1970] S. I. Pohožaev, "О собственных функциях квазилинейных эллиптических задач", *Mat. Sb.* (*N.S.*) **124**:2 (1970), 192–212. Translated as "On the eigenfunctions of quasilinear elliptic problems" in *Math. USSR Sb.* **11**:2 (1970), 171–188. MR 42 #8081 Zbl 0217.13203

[Rey 1992] O. Rey, "Concentration of solutions to elliptic equations with critical nonlinearity", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **9**:2 (1992), 201–218. MR 93e:35040 Zbl 0761.35034

[Servadei and Valdinoci $\geq$ 2014] R. Servadei and E. Valdinoci, "The Brézis–Nirenberg result for the fractional Laplacian", *Trans. Amer. Math. Soc.*. To appear.

[Sobolev 1938] S. L. Sobolev, "Об одной теореме функционального анализа", *Mat. Sb.* (*N.S.*) **46**:3 (1938), 471–497. Translated as "On a theorem of functional analysis" in *Eleven Papers on Analysis*, Amer. Math. Soc. Transl. (2) **34** (1963), 39–68. JFM 64.1100.02

[Stinga and Torrea 2010] P. R. Stinga and J. L. Torrea, "Extension problem and Harnack's inequality for some fractional operators", *Comm. Partial Differential Equations* **35**:11 (2010), 2092–2122. MR 2012c:35456 Zbl 1209.26013

[Tan 2011] J. Tan, "The Brézis–Nirenberg type problem involving the square root of the Laplacian", *Calc. Var. Partial Differential Equations* **42**:1-2 (2011), 21–41. MR 2012e:35079 Zbl 1248.35078

[Tarantello 1992] G. Tarantello, "On nonhomogeneous elliptic equations involving critical Sobolev exponent", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **9**:3 (1992), 281–304. MR 93i:35043 Zbl 0785.35046

Received April 11, 2013. Revised July 8, 2013.

EDUARDO COLORADO
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD CARLOS III DE MADRID
AVENIDA UNIVERSIDAD 30
28911 LEGANÉS
SPAIN

ecolorad@math.uc3m.es

and

INSTITUTO DE CIENCIAS MATEMÁTICAS, ICMAT (CSIC-UAM-UC3M-UCM)
C/ NICOLÁS CABRERA 15
28049 MADRID
SPAIN.

eduardo.colorado@icmat.es

ARTURO DE PABLO
DEPARTAMENTO DE MATEMÁTICAS
UNIVERSIDAD CARLOS III DE MADRID
AVENIDA UNIVERSIDAD 30
28911 LEGANÉS
SPAIN

arturop@math.uc3m.es

URKO SÁNCHEZ
UNIVERSIDAD CARLOS III DE MADRID
28911 LEGANÉS
SPAIN

urko.sanchez@gmail.com

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.87

A DENSITY THEOREM IN PARAMETRIZED DIFFERENTIAL GALOIS THEORY

THOMAS DREYFUS

We study parametrized linear differential equations with coefficients depending meromorphically upon the parameters. As a main result, analogously to the unparametrized density theorem of Ramis, we show that the parametrized monodromy, the parametrized exponential torus and the parametrized Stokes operators are topological generators in the Kolchin topology for the parametrized differential Galois group introduced by Cassidy and Singer. We prove an analogous result for the global parametrized differential Galois group, which generalizes a result by Mitschi and Singer. These authors give also a necessary condition on a group for being a global parametrized differential Galois group; as a corollary of the density theorem, we prove that their condition is also sufficient. As an application, we give a characterization of completely integrable equations, and we give a partial answer to a question of Sibuya about the transcendence properties of a given Stokes matrix. Moreover, using a parametrized Hukuhara–Turrittin theorem, we show that the Galois group descends to a smaller field, whose field of constants is not differentially closed.

Work partially supported by ANR, contract ANR-06-JCJC-0028.
MSC2010: 12H20, 34M15, 34M03.
Keywords: parametrized differential Galois theory, Stokes phenomenon.

Introduction

Let us consider a linear differential system of the form

$$\partial_z Y(z) = A(z)Y(z),$$

where $\partial_z = d/dz$, and $A(z)$ is an $m \times m$ matrix whose entries are germs of meromorphic functions in a neighborhood of a point, say 0 to fix ideas. The differential Galois group, which measures the algebraic dependencies among the solutions, can be viewed as an algebraic subgroup of $\mathrm{GL}_m(\mathbb{C})$ via the injective group morphism

$$\begin{aligned} \rho_U : \mathrm{Gal} &\to \mathrm{GL}_m(\mathbb{C}), \\ \sigma &\mapsto U(z)^{-1}\sigma(U(z)), \end{aligned}$$

where $U(z)$ is some arbitrary fundamental solution, i.e., an invertible solution matrix.

Let $U(z)$ be a fundamental solution contained in a Picard–Vessiot extension of the equation $\partial_z Y(z) = A(z)Y(z)$. The linear differential equation is said to be regular singular at 0 if there exists an invertible matrix $P(z)$ whose entries are germs of meromorphic functions such that $W(z) = P(z)U(z)$ satisfies

$$\partial_z W(z) = \frac{A_0}{z} W(z),$$

where A_0 is a matrix with constant complex entries. In this case, $W(z)$ usually involves multivalued functions. Analytic continuation of $W(z)$ along any simple loop γ around 0 yields another fundamental solution $W(z)M_\gamma$. The matrix M_γ, which is a monodromy matrix, has complex entries and depends only on the homotopy class of γ. The Schlesinger theorem says that the Zariski closure of the group generated by the monodromy matrix is the Galois group. In the general case, i.e., in the presence of an irregular singularity, the monodromy is no longer sufficient to provide a complete collection of topological generators. Ramis has shown that the group generated by the monodromy, the exponential torus and the Stokes operators, which is defined in a transcendental way as a subgroup of the differential Galois group, is dense in the latter in the Zariski topology.

More recently, a Galois theory for parametrized linear differential equations of the form

$$\partial_z Y(z,t) = A(z,t)Y(z,t), \tag{$*$}$$

where $t = (t_1, \dots, t_n)$ are parameters and A is a matrix whose entries lie in a certain field (specified explicitly throughout), has been developed in [Cassidy and Singer 2007] (henceforth abbreviated [CS]); see also [Hardouin and Singer 2008; Landesman 2008; Robinson 1959; Umemura 1996]. Namely, the Galois group, which measures the $(\partial_{t_1}, \dots, \partial_{t_n})$-differential and algebraic dependencies among the solutions, can be seen as a differential group in the sense of Kolchin, that is, a group of matrices whose entries lie in a differential field and satisfy a set of polynomial differential equations in the variables $t_1, \dots, t_n$; see [Cassidy 1972; 1989; Kolchin 1973; 1985; Minchenko and Ovchinnikov 2011]. The theory from [CS] requires the field of constants with respect to ∂_z to be of characteristic 0 and differentially closed (see Section 2A). The drawback of this latter assumption is that a differentially closed field is a very big field, and cannot be interpreted as a field of functions.

There is a link between the parametrized differential Galois theory and isomonodromy for equations with only regular singular poles (see [Cassidy and Singer 2007; Mitschi and Singer 2012; 2013]. Let

$$\mathcal{D}(t_0, r) = \{(z_1, \dots, z_n) \in \mathbb{C}^n \mid |z_i - t_{0,i}| < r \text{ for all } i \le n\}$$

be an open polydisc in $\mathbb{C}^n$, let $\mathcal{D}$ be an open subset of $\mathbb{C}$, and let $A(z,t)$ be a matrix whose entries are analytic on $\mathcal{D} \times \mathcal{D}(t_0, r)$. We consider open disks D_j that cover $\mathcal{D}$, and solutions $U_j(z,t)$ of $(*)$ that are analytic on $D_j \times \mathcal{D}(t_0, r)$. If $D_i \cap D_j \neq \varnothing$, we define the connection matrices $C_{i,j}(t) = U_i(z,t)^{-1}U_j(z,t)$. Following Definition 5.2 in [CS] (see also [Bolibruch 1997; Malgrange 1983]), the parametrized linear differential equation $(*)$ is said to be isomonodromic if there is a choice of (D_i) covering $\mathcal{D}$ and of the solutions $U_i(z,t)$ of $(*)$, analytic on $D_i \times \mathcal{D}(t_0, r)$, such that the connection matrices are independent of t. In this case, the matrix of the monodromy is constant on the polydisc $\mathcal{D}(t_0, r)$. When $A(z,t)$ is of the form $\sum_{i=1}^{s} A_i(t)/(z - u_i)$ such that all the $A_i(t)$ have analytic entries on U and $u_i \in \mathcal{D}$, the following statements are equivalent (see [CS], Propositions 5.3 and 5.4):

- The Galois group is conjugate over a differentially closed field (Definition 2.2) to a group of constant matrices.
- The parametrized linear differential equation is isomonodromic in the above sense.

- The parametrized linear differential equation is completely integrable (see Definition 3.1).

We are interested in the case where the parametrized linear differential equation may have irregular singularities, in a sense we are going to explain. The main result of this paper is a parametrized analogue of the density theorem of Ramis: we give topological generators for the Galois group in the Kolchin topology (in which closed sets are zero sets of differential algebraic polynomials). As an application of our main result, we improve Proposition 3.9 in [CS] (see Remark 3.4): a parametrized linear differential equation is completely integrable if and only if the topological generators for the Galois group just mentioned are conjugate to constant matrices over a field of meromorphic functions. Notice that the latter is not differentially closed.

The article is organized as follows. In the first section we study parametrized linear differential systems from an analytic point of view. The parameters will vary in U, a nonempty polydisc in $\mathbb{C}^n$. Let $t=(t_1,\dots,t_n)\in U$ denote the multiparameter. Let $\mathcal{M}_U$ be the field of meromorphic functions on U and let $\hat{K}_U=\mathcal{M}_U[[z]][z^{-1}]$. The Hukuhara–Turrittin theorem in this case gives the following result (see Remark 1.6 for a discussion of a similar result present in [Schäfke 2001]):

Proposition 1.3. *Consider the equation $\partial_z Y(z,t)=A(z,t)Y(z,t)$, with $A(z,t)\in \mathrm{M}_m(\hat{K}_U)$ (that is, an $m\times m$ matrix with entries in $\hat{K}_U$). Then there exist a nonempty polydisc $U'\subset U$ and $\nu\in\mathbb{N}^*$ such that we have a fundamental solution $F(z,t)$ of the form*

$$F(z,t)=\hat{H}(z,t)z^{L(t)}e^{Q(z,t)},$$

where:

- $\hat{H}(z,t)\in \mathrm{GL}_m(\hat{K}_{U'}[z^{1/\nu}])$.
- $L(t)\in \mathrm{M}_m(\mathcal{M}_{U'})$.
- $e^{Q(z,t)}=\mathrm{Diag}(e^{q_i(z,t)})$, *with* $q_i(z,t)\in z^{-1/\nu}\mathcal{M}_{U'}[z^{-1/\nu}]$.
- *Moreover*, *we have* $z^{L(t)}e^{Q(z,t)}=e^{Q(z,t)}z^{L(t)}$.

See Remark 1.4 for a discussion about the uniqueness of a fundamental solution of $(*)$ written in this way.

In Section 1C, we briefly review the Stokes phenomenon in the unparametrized case. We have solutions that are analytic in some sector and Gevrey asymptotic to the formal part of the solution in the Hukuhara–Turrittin canonical form. The fact that various asymptotic solutions do not glue to a single solution on the Riemann surface of the logarithm is called the Stokes phenomenon.

Let U be a nonempty polydisc in $\mathbb{C}^n$ and let $f(z,t)=\sum f_i(t)z^i\in\hat{K}_U$. We say that $f(z,t)$ belongs to $\mathcal{O}_U(\{z\})$ if for all $t\in U$, $z\mapsto\sum f_i(t)z^i$ is the germ of a meromorphic function at 0. Remark that if

$$f(z,t) \in \mathbb{O}_U(\{z\}) \subset \mathcal{M}_U[\![z]\!][z^{-1}] = \hat{K}_U,$$

then the z-coefficients $f_i(t)$ of $f(z,t)$ are analytic on U.

In Section 1D, we study the Stokes phenomenon for equations of the form $(*)$ with $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$. In particular, we prove that the asymptotic solutions depend analytically (under mild conditions) upon the parameters.

In the second section, we use the parametrized Hukuhara–Turrittin theorem to deduce some Galois-theoretic properties of parametrized linear differential equations in coefficients in $\mathbb{O}_U(\{z\})$. We first recall some facts from [CS] about parametrized differential Galois theory. The problem is that the theory in this reference cannot be applied here, since $\mathcal{M}_U$, our field of constants with respect to ∂_z, is a field of functions that are meromorphic in $t_1, \dots, t_n$, and this field is not differentially closed (see Section 2A). In the papers [Gillet et al. 2013; Wibmer 2012], the authors prove the existence of parametrized Picard–Vessiot extensions under weaker assumptions than in [CS]. See also [Chatzidakis et al. 2008; Peón Nieto 2011]. We do not use these latter results because we need a parametrized Hukuhara–Turrittin theorem (which proves directly that a parametrized Picard–Vessiot extension exists, not necessarily unique) in order to study the parametrized Stokes phenomenon. This allow us to define a group that we will call, by abuse of language, the parametrized differential Galois group; see Remark 2.8. In Section 2D we consider the local case of $(*)$, with $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$. We state and show the main result:

Theorem 2.20 (parametrized analogue of the density theorem of Ramis). *The group generated by the parametrized monodromy, the parametrized exponential torus and the parametrized Stokes operators is dense in the parametrized differential Galois group for the Kolchin topology.*

Then, we turn to the global case. We consider equations with coefficients in $\mathcal{M}_U(z)$ and study their global Galois group. We prove a density theorem in this global setting; see Theorem 2.24. The proof in the unparametrized case can be found in [Mitschi 1996]. In Section 2F, we give various examples of calculations.

In the third section, we give three applications. First, we prove a criterion for the integrability of differential systems (see Definition 3.1):

Proposition 3.2. *Let $A(z,t) \in \mathrm{M}_m(\mathcal{M}_U(z))$. Then the linear differential equation $\partial_z Y(z,t) = A(z,t)Y(z,t)$ is completely integrable if and only if there exists a fundamental solution such that the matrices of the parametrized monodromy, the parametrized exponential torus and the parametrized Stokes operators for all the singularities are constant, i.e., do not depend on z.*

As a second application, we give a partial answer to a question of Sibuya [1975] regarding the differential transcendence properties of a Stokes matrix of the parametrized linear differential equation

$$\begin{pmatrix} \partial_z Y(z,t) \\ \partial_z^2 Y(z,t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ z^3+t & 0 \end{pmatrix} \begin{pmatrix} Y(z,t) \\ \partial_z Y(z,t) \end{pmatrix}.$$

Sibuya was asking whether an entry of a given Stokes matrix at infinity is ∂_t-differentially transcendental, i.e., satisfies no differential polynomial equation. We prove that it is at least not ∂_t-finite, i.e., that it satisfies no linear differential equation.

As a last application, we deal with the inverse problem. We prove that if G is the global parametrized differential Galois group of some equation having coefficients in $k(z)$ (see Section 3C), then G contains a finitely generated Kolchin-dense subgroup. The converse of this latter assertion has been proved in Corollary 5.2 of [Mitschi and Singer 2012], and we obtain a result on the inverse problem:

Theorem 3.11. *G is the global parametrized differential Galois group of some equation having coefficients in $k(z)$ if and only if G contains a finitely generated Kolchin-dense subgroup.*

In the Appendix, we prove the following result:

Theorem A.1. *Consider the equation $\partial_z Y(z,t) = A(z,t)Y(z,t)$, with $A(z,t) \in \mathrm{M}_m(\hat{K}_U)$. Then there exists a nonempty polydisc $U' \subset U$ such that we have a fundamental solution $F(z,t)$ of the form*

$$F(z,t) = \hat{P}(z,t) z^{C(t)} e^{Q(z,t)},$$

where:

- $\hat{P}(z,t) \in \mathrm{GL}_m(\hat{K}_{U'})$,
- $C(t) \in \mathrm{M}_m(\mathcal{M}_{U'})$,
- $e^{Q(z,t)} = \mathrm{Diag}(e^{q_i(z,t)})$, *with* $q_i(z,t) \in z^{-1/\nu}\mathcal{M}_{U'}[z^{-1/\nu}]$, *for some* $\nu \in \mathbb{N}^*$.

Remark that contrary to Proposition 1.3, the entries of the formal part are not ramified. On the other hand, $z^{C(t)}$ and $e^{Q(z,t)}$ do not commute anymore. This theorem is not necessary for the proof of the main result of the paper; this is the reason why we give the proof in the Appendix. However, this result is important since it permits one to determine the equivalence classes (see [van der Put and Singer 2003, p. 7]) of parametrized linear differential systems in coefficients in $\hat{K}_U$.

1. Local analytic linear differential systems depending upon parameters

In Section 1A, we define the field to which the entries of the fundamental solution, in the Hukuhara–Turrittin canonical form, will belong. In Section 1B, we prove a parametrized version of the Hukuhara–Turrittin theorem. In Section 1C, we briefly review the Stokes phenomenon in the unparametrized case. In Section 1D, we study the Stokes phenomenon in the parametrized case.

1A. ***Definition of the fields.*** Let us consider a linear differential system of the form $\partial_z Y(z) = A(z)Y(z)$, where $A(z)$ is an $m \times m$ matrix whose entries belongs to $\mathbb{C}[\![z]\!][z^{-1}]$. We know we can find a formal fundamental solution in the Hukuhara–Turrittin canonical form $\hat{H}(z)z^L e^{Q(z)}$, where:

- $\hat{H}(z)$ is a matrix of formal power series in $z^{1/\nu}$ for some $\nu \in \mathbb{N}^*$.
- $L \in \mathrm{M}_m(\mathbb{C})$.
- $Q(z) = \mathrm{Diag}(q_i(z))$, with $q_i(z) \in z^{-1/\nu}\mathbb{C}[z^{-1/\nu}]$.
- Moreover, we have $z^L e^{Q(z)} = e^{Q(z)} z^L$.

Notice that this formulation is trivially equivalent to Theorem 3.1 in [van der Put and Singer 2003]. Let U be a nonempty polydisc of $\mathbb{C}^n$, and define $\hat{K}_U$ and $\mathcal{M}_U$ as on page 90. We want to construct a field containing a fundamental set of solutions of $(*)$, where $A(z,t) \in \mathrm{M}_m(\hat{K}_U)$. Let $\Delta_t = \{\partial_{t_1}, \dots, \partial_{t_n}\}$ and let

$$\mathbf{E}_U = \bigcup_{\nu \in \mathbb{N}^*} z^{-1/\nu}\mathcal{M}_U[z^{-1/\nu}].$$

We define formally the (∂_z, Δ_t)-ring, i.e., a ring equipped with $n+1$ derivations $\partial_z, \partial_{t_1}, \dots, \partial_{t_n}$, a priori not required to commute with each other, to be

$$R_U := \hat{K}_U\big[\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U}, l(e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}\big],$$

with the following rules:

(1) The symbols log, $(z^{a(t)})_{a(t)\in\mathcal{M}_U}$ and $(e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}$ only satisfy the following relations:

$$z^{a(t)+b(t)} = z^{a(t)}z^{b(t)}, \quad z^a = z^a \in \hat{K}_U \quad \text{for } a \in \mathbb{Z},$$
$$e(q_1(z,t)+q_2(z,t)) = e(q_1(z,t))e(q_2(z,t)), \quad e(0) = 1.$$

(2) The following rules of differentiation:

$$\partial_z \log = z^{-1}, \quad \partial_{t_i} \log = 0, \quad \partial_z z^{a(t)} = \frac{a(t)}{z} z^{a(t)}, \quad \partial_{t_i} z^{a(t)} = \partial_{t_i}(a(t)) \log z^{a(t)},$$
$$\partial_z e(q(z,t)) = \partial_z(q(z,t))e(q(z,t)), \quad \partial_{t_i} e(q(z,t)) = \partial_{t_i}(q(z,t))e(q(z,t)),$$

equip the ring with a (∂_z, Δ_t)-differential structure, since these rules descend to the quotient, as can be readily checked.

The intuitive interpretations of these symbols are: $\log = \log(z)$, $z^{a(t)} = e^{a(t)\log(z)}$ and $e(q(z,t)) = e^{q(z,t)}$. Let $f(z,t)$ be one these latter functions. Then $f(z,t)$ has a natural interpretation as an analytic function on $\widetilde{\mathbb{C}} \times U'$, where $\widetilde{\mathbb{C}}$ is the Riemann surface of the logarithm and U' is some nonempty polydisc contained in U. We will use the analytic function instead of the symbol when we will consider asymptotic solutions (see Section 1C and Section 1D). For the time being, however, we see them only as symbols.

Let $\overline{\mathcal{M}}_U$ be the algebraic closure of $\mathcal{M}_U$. In the same way as for R_U, we construct the (∂_z, Δ_t)-ring

$$\overline{R}_U := \overline{\mathcal{M}}_U[\![z]\!][z^{-1}]\big[\log, (z^{a(t)})_{a(t)\in\overline{\mathcal{M}}_U}, (e(q(z,t)))_{q(z,t)\in\mathscr{E}_U}\big],$$

where

$$\mathscr{E}_U = \bigcup_{\nu\in\mathbb{N}^*} z^{-1/\nu}\overline{\mathcal{M}}_U[z^{-1/\nu}].$$

and its field of fractions has field of constants with respect to ∂_z equal to $\overline{\mathcal{M}}_U$. Since $R_U \subset \overline{R}_U$, R_U is also an integral domain. Therefore, we may consider the (∂_z, Δ_t)-fields

$$K_{F,U} = \mathcal{M}_U(\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U}),$$
$$\hat{K}_{F,U} = \hat{K}_U(\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U}),$$

and

$$(\mathbf{K}_U)^\wedge = \hat{K}_U\big(\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U}, (e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}\big).$$

In the definition of the fields $K_{F,U}$ and $\hat{K}_{F,U}$, the subscript F stands for Fuchsian. Since $(\mathbf{K}_U)^\wedge$ is contained in the field of fractions of $\overline{R}_U$, its field of constants with respect to ∂_z is equal to $\overline{\mathcal{M}}_U \cap (\mathbf{K}_U)^\wedge = \mathcal{M}_U$.

We have defined (∂_z, Δ_t)-fields where all the derivations commute with each other. We have the following inclusions of (∂_z, Δ_t)-fields:

$$\begin{array}{ccccccc} & & K_{F,U} & & & & \\ & \nearrow & & \searrow & & & \\ \mathcal{M}_U & \to & \hat{K}_U & \to & \hat{K}_{F,U} & \to & (\mathbf{K}_U)^\wedge. \end{array}$$

Remark 1.1. Any algebraic function over $\mathcal{M}_U$ can be seen as an element of $\mathcal{M}_{U'}$ for some nonempty $U' \subset U$. Therefore, a finite extension of $\mathcal{M}_U$ can be embedded in $\mathcal{M}_{U'}$ for a convenient choice of $U' \subset U$. We will use this fact in the rest of the paper.

Lemma 1.2. *Let $U \subseteq \mathbb{C}^n$ be a nonempty polydisc, and let $L(t) \in \mathrm{M}_m(\overline{\mathcal{M}}_U)$, where $\overline{\mathcal{M}}_U$ is the algebraic closure of $\mathcal{M}_U$. There exist a nonempty polydisc $U' \subset U$ and $z^{L(t)} \in \mathrm{GL}_m(K_{F,U'})$ satisfying*

$$\partial_z z^{L(t)} = \frac{L(t)}{z} z^{L(t)} = z^{L(t)} \frac{L(t)}{z}.$$

Proof. Let

$$L(t) = P(t)(D(t) + N(t))P^{-1}(t)$$

be the Jordan decomposition of $L(t)$, where $D(t) = \mathrm{Diag}(d_i(t))$ with $d_i(t) \in \overline{\mathcal{M}}_U$, $N(t)$ is nilpotent, $D(t)N(t) = N(t)D(t)$ and $P(t) \in \mathrm{GL}_m(\overline{\mathcal{M}}_U)$.

Due to Remark 1.1, there exists a nonempty polydisc $U' \subset U$ such that $d_i(t) \in \mathcal{M}_{U'}$ and $P(t) \in \mathrm{GL}_m(\mathcal{M}_{U'})$. We may restrict U' and assume that $N(t)$ does not depend upon t in U'. Let us write $N := N(t)$. Then the matrix

$$z^{L(t)} = P(t)\,\mathrm{Diag}(z^{d_i(t)})e^{N\log}P^{-1}(t)$$

belongs to $\mathrm{GL}_m(K_{F,U'})$, and $z^{L(t)}$ satisfies

$$\partial_z z^{L(t)} = \frac{L(t)}{z} z^{L(t)} = z^{L(t)}\frac{L(t)}{z}. \qquad \square$$

Let $a(t) \in \mathcal{M}_U$ and let $(a(t)) \in \mathrm{M}_1(\mathcal{M}_U)$ be the corresponding matrix. Then we have $z^{a(t)} = z^{(a(t))}$.

1B. ***The Hukuhara–Turrittin theorem in the parametrized case.*** The goal of this subsection is to give the parametrized version of the Hukuhara–Turrittin theorem. In the Appendix, we prove a slightly different result, which is not needed in the paper; see Theorem A.1.

Proposition 1.3. *Let U be a nonempty polydisc in $\mathbb{C}^n$ and consider the equation*

$$\partial_z Y(z,t) = A(z,t)Y(z,t),$$

with $A(z,t) \in \mathrm{M}_m(\hat{K}_U)$. There exists a nonempty polydisc $U' \subset U$ such that we have a fundamental solution $F(z,t) \in \mathrm{GL}_m((\mathbf{K}_{U'})^\wedge)$ of the form

$$F(z,t) = \hat{H}(z,t)z^{L(t)}e(Q(z,t)),$$

where:

- $\hat{H}(z,t) \in \mathrm{GL}_m(\hat{K}_{U'}[z^{1/\nu}])$, *for some $\nu \in \mathbb{N}^*$.*
- $L(t) \in \mathrm{M}_m(\mathcal{M}_{U'})$.
- $e(Q(z,t)) = \mathrm{Diag}(e(q_i(z,t)))$, *with $q_i(z,t) \in \mathbf{E}_{U'}$.*
- *Moreover, we have $e(Q(z,t))z^{L(t)} = z^{L(t)}e(Q(z,t))$.*

Furthermore, if $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$, there exists a nonempty polydisc $U'' \subset U'$ such that we may assume that the z-coefficients of $\hat{H}(z,t)$ are all analytic on U''.

Remark 1.4. Remark that we have no uniqueness of the fundamental solution written in this way, since $z^{\kappa}\hat{H}(z,t)z^{L(t)-\kappa}e^{Q(z,t)}$ is also a fundamental solution for all $\kappa \in \mathbb{Z}$. However, by the construction of $(\mathbf{K}_{U'})^\wedge$, if $\hat{H}_i(z,t)z^{L_i(t)}e(Q_i(z,t))$ are fundamental solutions of $(*)$ written in this way for $i = 1, 2$, then, up to a permutation, Q_1 and Q_2 have the same entries.

Example 1.5 [Schäfke 2001, Introduction]. If we consider

$$z^2\partial_z Y(z,t) = \begin{pmatrix} t & 1 \\ z & 0 \end{pmatrix} Y(z,t),$$

we get the solution

$$\left(\begin{pmatrix} 1 & 1 \\ 0 & -t \end{pmatrix} + O(z)\right)\begin{pmatrix} z^{1/t}e^{-t/z} & 0 \\ 0 & z^{-1/t} \end{pmatrix} \tag{1-1}$$

for $t \neq 0$, and the solution

$$\begin{pmatrix} 1 & 1 \\ z^{1/2} & -z^{1/2} \end{pmatrix}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + O(z^{1/2})\right)\begin{pmatrix} z^{1/4}e^{-z^{-1/2}} & 0 \\ 0 & z^{1/4}e^{z^{-1/2}} \end{pmatrix}$$

for $t = 0$. The latter is not the specialization of (1-1) at $t = 0$. The problem is that the level of the unparametrized system (see Section 1C for the definition) at $t = 0$ is 1 and the level of the unparametrized system for $t \neq 0$ is $\frac{1}{2}$. This example shows that we cannot get a solution in the parametrized Hukuhara–Turrittin form that remains valid for all values of the parameter t. This is the reason why we have to restrict the subset of the parameter space.

Remark 1.6. Similar results to Proposition 1.3 have been proved in Theorem 4.2 of [Schäfke 2001]. We now explain the result of Schäfke. Let U be an open connected subset of $\mathbb{C}^n$ that contains 0, and let $A(z,t) = \sum_{l=s}^{\infty} A_l(t)$, with $s \in \mathbb{Z}$ and $A_l(t)$ analytic in U. In particular, $A(z,t) \in \mathrm{M}_m(\hat{K}_U)$. Assume that, for all $t \in U$, there exists a solution $\hat{H}_t(z)z^{L_t}e(Q(z,t))$ to $(*)$ given in the classical Hukuhara–Turrittin canonical form, i.e., such that:

- The z-coefficients of the $q_i(z,t)$ are analytic functions in $t \in U$.
- The degree in z^{-1} of $q_i(z,t) - q_j(z,t)$ is independent of t in U.
- If $q_i(z,t) \not\equiv q_j(z,t)$, then $q_i(z,0) \neq q_0(z,0)$.

Under these assumptions, Schäfke concludes that there exists an open neighborhood $U' \subset U$ of 0 in the t-plane such that there exists a solution

$$\hat{H}(z,t)z^{L(t)}e(Q(z,t)) \in \mathrm{GL}_m((\mathbf{K}_{U'})^\wedge)$$

with $\hat{H}(z,t) = \sum_{l=0}^{\infty} \hat{H}_l(t)$ and such that the maps $t \mapsto \hat{H}_l(t)$, $L(t)$ are analytic. Notice that Schäfke gives a necessary and sufficient condition, that can be algorithmically checked, for well-behaved exponential part. See [ibid., Theorem 5.2]. Using Schäfke's theorem, we can deduce Proposition 1.3 only in the particular case where $A(z,t)$ has entries with z-coefficients analytic in U. Note that [ibid.] does not allow us to deduce the general case. See also [Babbitt and Varadarajan 1985, §10, Theorem 1] for another result of this nature.

Proof of Proposition 1.3. Let $K = C[\![z]\!][z^{-1}]$, where C is an algebraically closed field of characteristic 0, equipped with a derivation ∂_z that acts trivially on C and with $\partial_z(z) = 1$. The Hukuhara–Turrittin theorem (see Theorem 3.1 in [van der Put and Singer 2003]) is valid for linear differential system with entries in K. We apply it with $C = \overline{\mathcal{M}}_U$, the algebraic closure of $\mathcal{M}_U$.

Let us consider the matrices $L(t) \in \mathrm{M}_m(\overline{\mathcal{M}}_U)$ and $Q(z,t) = \mathrm{Diag}(q_i(z,t))$, with $q_i(z,t) \in z^{-1/\nu}\overline{\mathcal{M}}_U[z^{-1/\nu}]$ for some $\nu \in \mathbb{N}$. Because of Remark 1.1 and Lemma 1.2, there exists a nonempty polydisc $U' \subset U$ such that we may define $z^{L(t)} \in \mathrm{GL}_m(K_{F,U'})$ satisfying

$$\partial_z z^{L(t)} = \frac{L(t)}{z} z^{L(t)} = z^{L(t)} \frac{L(t)}{z},$$

$L(t) \in \mathrm{M}_m(\mathcal{M}_{U'})$ and $q_i(z,t) \in \mathbf{E}_{U'}$. Hence, the Hukuhara–Turrittin theorem gives a fundamental solution

$$F'(z,t) = \hat{H}'(z,t) z^{L(t)} e(Q(z,t))$$

on U', where:

- $\hat{H}'(z,t) \in \mathrm{GL}_m(\overline{\mathcal{M}}_{U'}[\![z^{1/\nu}]\!][z^{-1/\nu}])$, for some $\nu \in \mathbb{N}$.
- $L(t) \in \mathrm{M}_m(\mathcal{M}_{U'})$.
- $e(Q(z,t)) = \mathrm{Diag}(e(q_i(z,t)))$, with $q_i(z,t) \in \mathbf{E}_{U'}$.
- Moreover, we have $e(Q(z,t)) z^{L(t)} = z^{L(t)} e(Q(z,t))$.

Let us prove now that we may find $\hat{H}(z,t) \in \mathrm{GL}_m(\hat{K}_{U'}[z^{1/\nu}])$ such that

$$F(z,t) = \hat{H}(z,t) z^{L(t)} e(Q(z,t))$$

is a fundamental solution. The matrix

$$F'(z,t) = \hat{H}'(z,t) z^{L(t)} e(Q(z,t))$$

satisfies the parametrized linear differential equation

$$\partial_z F'(z,t) = A(z,t) F'(z,t),$$

and the matrix $z^{L(t)} e(Q(z,t))$ satisfies the parametrized linear differential equation

$$\begin{aligned} \partial_z z^{L(t)} e(Q(z,t)) &= \big(z^{-1} L(t) + \partial_z Q(z,t)\big) z^{L(t)} e(Q(z,t)) \\ &= z^{L(t)} e(Q(z,t)) \big(z^{-1} L(t) + \partial_z Q(z,t)\big). \end{aligned}$$

Hence

$$\partial_z \hat{H}'(z,t) = A(z,t) \hat{H}'(z,t) - \hat{H}'(z,t)\big(z^{-1} L(t) + \partial_z Q(z,t)\big).$$

We write $\hat{H}'(z,t)$ as a column vector $\widetilde{H}'(z,t)$ of size m^2. Let

$$C(z,t) \in \mathrm{M}_{m^2}(\hat{K}_{U'}[z^{1/\nu}]),$$

with $\nu \in \mathbb{N}^*$ such that $\widetilde{H}'(z,t)$ satisfies the parametrized linear differential system

$$\partial_z \widetilde{H}'(z,t) = C(z,t)\widetilde{H}'(z,t).$$

Let us write

$$\widetilde{H}'(z,t) = \sum_{i \geq N} \widetilde{H}'_i(t) z^{i/\nu} \quad \text{and} \quad C(z,t) = \sum_{i \geq M} C_i(t) z^{i/\nu},$$

where $M, N \in \mathbb{Z}$. Then, by identifying the coefficients of the $z^{i/\nu}$-terms of the power series in the equation $\partial_z \widetilde{H}'(z,t) = C(z,t)\widetilde{H}'(z,t)$, we find that

$$\left(\frac{i}{\nu}+1\right)\widetilde{H}'_{i+\nu}(t) = \sum_{l=N}^{i-M} C_{i-l}(t)\widetilde{H}'_l(t).$$

By the definition of $\hat{K}_{U'}[z^{1/\nu}]$, every $C_i(t)$ belongs to $\mathrm{M}_m(\mathcal{M}_{U'})$. The fact that there exists a fundamental solution $\hat{H}(z,t)z^{L(t)}e(Q(z,t))$ with $\hat{H}(z,t) \in \mathrm{GL}_m(\hat{K}_{U'}[z^{1/\nu}])$ is now clear.

Assume now that $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$. Let U'' be a nonempty polydisc with $U'' \subset U'$ such that for $z \neq 0$ fixed, the entries of the z-coefficients of

$$z^{-1}L(t) + \partial_z Q(z,t)$$

are analytic on U''. Then the entries of the z-coefficients of $C(z,t)$ are all analytic on U''. Hence, we may assume that the entries of the z-coefficients of $\hat{H}(z,t)$ are all analytic on U''. □

Remark 1.7. If we take a smaller nonempty polydisc U, we may assume that if we consider the equation $(*)$ with $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$, then the fundamental solution of Proposition 1.3 belongs to $\mathrm{GL}_m((\mathbf{K}_U)^\wedge)$, and the entries of the z-coefficients of $\hat{H}(z,t)$ are all analytic on U.

1C. ***Review of the Stokes phenomenon in the unparametrized case.*** In this subsection we will briefly review the Stokes phenomenon in the unparametrized case. See [Cano and Ramis 1995; Écalle 1981; Loday-Richaud 1990; 1994; 1995; Loday-Richaud and Remy 2011; Malgrange 1991; 1995; Malgrange and Ramis 1992; Ramis 1980; 1985; Rasoamanana 2010; Remy 2012; Ramis and Sibuya 1989; Singer 2009; Wasow 1965], and in particular Chapter 8 of [van der Put and Singer 2003] for more details. We will generalize some results concerning the summation of divergent series in the parametrized case in Section 1D. First we treat the example of the Euler equation

$$z^2\partial_z Y(z) + Y(z) = z,$$

which admits as a solution the formal series $\hat{f}(z)=\sum_{n=0}^{\infty}(-1)^n n!z^{n+1}$. Classical methods of differential equations give another solution:

$$f(z)=\int_0^z e^{1/z}e^{-1/t}\frac{dt}{t}=\int_0^{\infty}\frac{1}{1+u}e^{-u/z}\,du,$$

where $1/t-1/z=u/z$. The solution $\hat{f}(z)$ is divergent and the solution $f(z)$ can be extended to an analytic function on the sector $V=\Sigma(-3\pi/2,3\pi/2)$, where, here and throughout, we use the notation

$$\Sigma(\alpha,\beta):=\{z\in\widetilde{\mathbb{C}}\mid \arg(z)\in\,]\alpha,\beta[\}$$

to represent sectors in $\widetilde{\mathbb{C}}$. On this sector, $f(z)$ is 1-Gevrey asymptotic to $\hat{f}(z)$: for every closed subsector W of V, there exist $A_W\in\mathbb{R}$ and $\varepsilon>0$ such that for all N and all $z\in W$ with $|z|<\varepsilon$,

$$\left|f(z)-\sum_{n=0}^{N-1}(-1)^n n!z^{n+1}\right|\le(A_W)^{N+1}(N+1)!|z|^{N+1}.$$

We can also consider $f(e^{2i\pi}z)$, which is an asymptotic solution on the sector

$$V'=\Sigma(\pi/2,7\pi/2).$$

The two asymptotic solutions do not glue to a single asymptotic solution on $V\cup V'$. In fact, the residue theorem implies that the difference in $V\cap V'$ of the two asymptotic solutions is

$$2i\pi e^{1/z}.$$

The fact that various asymptotic solutions do not glue to a single analytic solution is called the Stokes phenomenon.

More generally, let us consider a linear differential equation $\partial_z Y(z)=A(z)Y(z)$ such that the entries of $A(z)$ are germs of meromorphic functions in a neighborhood of 0. Let $\hat{H}(z)z^L e(Q(z))$ be a fundamental solution in the Hukuhara–Turrittin canonical form, with $Q(z)=\mathrm{Diag}(q_i(z))$. Since

$$\hat{H}(z)z^L e(Q(z))=\hat{H}(z)\,\mathrm{Diag}(z^k)z^{L-k\,\mathrm{Id}}e(Q(z))$$

for all $k\in\mathbb{N}$, we may assume that $\hat{H}(z)$ has no pole at $z=0$. The levels of $\partial_z Y(z)=A(z)Y(z)$ are the degrees in z^{-1} of the $q_i(z)-q_j(z)$ (the levels are positive rational numbers and are well-defined because of Remark 1.4). Consider

$$q(z)=q_k z^{-k/\nu}+\cdots+q_1 z^{-1/\nu}\in z^{-1/\nu}\mathbb{C}[z^{-1/\nu}]$$

with $\nu\in\mathbb{N}$. The real number d is called singular for $q(z)$ if $q_k e^{-idk/\nu}$ is a positive real number. These correspond to the arguments d such that $r\mapsto e^{q(re^{id})}$ increases fastest as r tends to 0^+. The singular directions of $\partial_z Y(z)=A(z)Y(z)$ (we will write singular directions when no confusion is likely to arise) are the real numbers that are singular for one of the $q_i(z) - q_j(z)$, with $i \neq j$. Notice that the set of singular directions is finite modulo $2\pi\nu$ for some $\nu \in \mathbb{N}$. Let $k_1 < \cdots < k_r$ be the levels of the linear differential equation. There exists a decomposition $\hat{H}(z) = \hat{H}_{k_1}(z) + \cdots + \hat{H}_{k_r}(z)$ such that for d not a singular direction, there exists an unique r-tuple of matrices $(H^d_{k_1}(z), \ldots, H^d_{k_r}(z))$ such that $H^d_{k_i}(z)$ is analytic on the sector

$$V_d = \Sigma(d - \pi/2k_i, d + \pi/2k_i),$$

and is k_i-Gevrey asymptotic to $\hat{H}_{k_i}(z) = \sum_{n\in\mathbb{N}} \hat{H}_{n,k_i} z^n$ on V_d: for every closed subsector W of V_d, there exist $A_W \in \mathbb{R}$ and $\varepsilon > 0$ such that for all N and all $z \in W$ with $|z| < \varepsilon$,

$$\left| H^d_{k_i}(z) - \sum_{n=0}^{N-1} \hat{H}_{n,k_i} z^n \right| \le (A_W)^N \Gamma\Big(1 + \frac{N}{k_i}\Big) |z|^N,$$

where Γ denotes the gamma function. Until the end of the paper, we will denote a fixed branch of the complex logarithm by $\log(z)$. Furthermore, the matrix

$$(H^d_{k_1}(z) + \cdots + H^d_{k_r}(z)) e^{L\log(z)} e^{Q(z)} = H^d(z) e^{L\log(z)} e^{Q(z)}, \tag{1-2}$$

which is analytic on the sector $\Sigma(d - \pi/2k_r, d + \pi/2k_r)$, is a solution of $\partial_z Y(z) = A(z)Y(z)$. As a matter of fact, $H^d_{k_i}(z)$ is k_i-Gevrey asymptotic to $\hat{H}_{k_i}(z)$ on the larger sector

$$\Sigma(d_l - \pi/2k_i, d_{l+1} + \pi/2k_i),$$

where d_l, d_{l+1} are two singular directions such that $]d_l, d_{l+1}[$ contains no singular directions. Therefore, we can construct an analytic solution on the sector $\Sigma(d_l - \pi/2k_r, d_{l+1} + \pi/2k_r)$. Let $d \in \mathbb{R}$, and choose $d^{\pm}$ such that

$$d - \frac{\pi}{2k_r} < d^- < d < d^+ < d + \frac{\pi}{2k_r}$$

and such that there are no singular directions in $[d^-, d[\,\cup\,]d, d^+]$. We get two matrices, $H^{d^+}(z) e^{L\log(z)} e^{Q(z)}$ and $H^{d^-}(z) e^{L\log(z)} e^{Q(z)}$, which are germs of analytic solutions on the sectors

$$\Sigma(d^- - \pi/2k_r, d + \pi/2k_r) \quad \text{and} \quad \Sigma(d - \pi/2k_r, d^+ + \pi/2k_r),$$

respectively. The two matrices are, in particular, germs of solutions of $\partial_z Y(z) = A(z)Y(z)$ on the sector

$$\Sigma(d - \pi/2k_r, d + \pi/2k_r).$$

A computation shows that there exists a matrix $\mathrm{St}_d \in \mathrm{GL}_m(\mathbb{C})$, which we call the Stokes matrix in the direction d, such that

$$H^{d^+}(z)e^{L\log(z)}e^{Q(z)} = H^{d^-}(z)e^{L\log(z)}e^{Q(z)}\,\mathrm{St}_d\,.$$

Proposition 1.8. *The following statements are equivalent*:

(1) *The entries of* $\hat{H}(z)$ *converge.*

(2) $\mathrm{St}_d = \mathrm{Id}$ *for all* $d \in \mathbb{R}$.

(3) $\mathrm{St}_d = \mathrm{Id}$ *for all singular directions.*

Proof. From what is preceding, we deduce that if d is not a singular direction, then $\mathrm{St}_d = \mathrm{Id}$. Therefore, the statements (2) and (3) are equivalent. If the entries of $\hat{H}(z)$ converge, then, since $\hat{H}(z)$ is Gevrey asymptotic to itself on every sector of $\widetilde{\mathbb{C}}$, $H^d(z) = \hat{H}(z)$ for all $d \in \mathbb{R}$, and (2) holds. Assume now that $\mathrm{St}_d = \mathrm{Id}$ for all singular directions. From the proof of [van der Put and Singer 2003, Theorem 8.10], we obtain that the entries of $\hat{H}(z)$ converge. □

We can compute the asymptotic solutions using the Laplace and the Borel transformations. See Chapters 2 and 3 of [Balser 1994] for more details.

Definition 1.9. (1) Let $k \in \mathbb{Q}$. The formal Borel transform $\hat{\mathcal{B}}_k$ is the map that transforms the formal power series $\sum a_n z^n$ into the formal power series

$$\hat{\mathcal{B}}_k\Big(\sum a_n z^n\Big) = \sum \frac{a_n}{\Gamma(1+n/k)} z^n.$$

(2) Let $d \in \mathbb{R}$, $k \in \mathbb{Q}$, $\varepsilon > 0$ and let f be analytic on the sector $\Sigma(d-\varepsilon, d+\varepsilon)$. We assume that there exist $A, B > 0$ such that

$$|f(z)| \le Ae^{B|z|^k}$$

for $\arg(z) = d$. Then the following integral is the germ of an analytic function on $\Sigma(d-\pi/2k, d+\pi/2k)$ (see [ibid., p. 13], for a proof), and is called the Laplace transform of order k in the direction d of f:

$$\mathcal{L}_{k,d}(f)(z) = \int_0^{\infty e^{id}} f(u)e^{-(u/z)^k}\, d\Big(\Big(\frac{u}{z}\Big)^k\Big).$$

For a proof of the following proposition, see Section 7.2 of [ibid.].

Proposition 1.10. *Let* $k_1 < \cdots < k_r$ *be the levels of* $\partial_z Y(z) = A(z)Y(z)$ *and set* $k_{r+1} = +\infty$. *Suppose that* $d \in \mathbb{R}$ *is not a singular direction, and let* $\hat{h}(z)$ *be an entry of* $\hat{H}(z)$. *Define* $(\kappa_1, \dots, \kappa_r)$ *by*

$$\kappa_i^{-1} = k_i^{-1} - k_{i+1}^{-1}.$$

The series $\hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}(\hat{h})$ *converges, and there exist* $\varepsilon_1, A_1, B_1 > 0$ *such that it has an analytic continuation* h_1 *on the sector* $\Sigma(d-\varepsilon_1, d+\varepsilon_1)$, *and*

$$|h_1(z)| \leq A_1 e^{B_1|z|^{\kappa_1}}$$

in this sector. Moreover, for $j = 2, \dots, r$ *there exist* $\varepsilon_j, A_j, B_j > 0$ *such that the function* $h_{j+1} = \mathcal{L}_{\kappa_j,d}(h_j)$ *is analytic on the sector* $\Sigma(d-\varepsilon_j, d+\varepsilon_j)$, *and*

$$|h_j(z)| \leq A_j e^{B_j|z|^{\kappa_j}}$$

on this sector. Therefore, we may apply $\mathcal{L}_{\kappa_r,d} \circ \cdots \circ \mathcal{L}_{\kappa_1,d} \circ \hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}$ *to every entry of* $\hat{H}(z)$. *We have the following equality*:

$$H^d(z) = \mathcal{L}_{\kappa_r,d} \circ \cdots \circ \mathcal{L}_{\kappa_1,d} \circ \hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}(\hat{H}).$$

1D. ***Stokes phenomenon in the parametrized case.*** Consider the equation $(*)$, with $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$ (see page 90), where U is a nonempty polydisc in $\mathbb{C}^n$, and consider $F(z,t) = \hat{H}(z,t)z^{L(t)}e(Q(z,t))$, with $Q(z,t) = \mathrm{Diag}(q_i(z,t))$, the fundamental solution of Proposition 1.3. Since for all $k \in \mathbb{N}$, $F(z,t)$ is equal to $\hat{H}(z,t)\,\mathrm{Diag}(z^k)z^{L(t)-k\,\mathrm{Id}}e(Q(z,t))$, we may assume that $\hat{H}(z,t)$ has no pole at $z=0$. We define the levels of the system $(*)$ as the levels of the specialized system. The levels may depend upon t, but they are invariant on the complement of a closed set with empty interior. We want to extend the definition of the singular directions to the parametrized case. Consider $q(z,t) = q_k(t)z^{-k/\nu} + \cdots + q_1(t)z^{-1/\nu} \in \mathbf{E}_U$. A continuous function $d : U \to \mathbb{R}$ is called singular for $q(z,t)$ if

$$q_k(t)e^{-id(t)k/\nu} \in \mathbb{R}^{\geq 0} \quad \text{for all } t \in U.$$

In general, the positive number $q_k(t)e^{-id(t)k/\nu}$ depends on t if $d(t)$ is a singular direction for $q(z,t)$. The singular directions of $(*)$ (we will just write singular directions when no confusion is likely to arise) are the directions that are singular for one of the $q_i(z,t) - q_j(z,t)$, with $i \neq j$.

Remark 1.11. (1) It may happen that for some $t_0 \in U$, the singular directions of $(*)$ evaluated at t_0 are not equal to the singular directions of the specialized system $\partial_z Y(z,t_0) = A(z,t_0)Y(z,t_0)$. Take for example $n=1$, $U=\mathbb{C}$, $t_0=0$ and $A(z,t) = \mathrm{Diag}(-2tz^{-3} - z^{-2}, 2tz^{-3} + z^{-2})$. The two exponentials are

$$e(q_1(z,t)) = e(tz^{-2} + z^{-1}) \quad \text{and} \quad e(q_2(z,t)) = e(-tz^{-2} - z^{-1}).$$

However, there exists $V \subset U$, a closed set with empty interior, such that for all t_0 in $U \setminus V$, the singular directions of $(*)$ evaluated at t_0 are equal to the singular directions of the specialized system $\partial_z Y(z,t_0) = A(z,t_0)Y(z,t_0)$.

(2) Unfortunately, two different singular directions may be equal on a subset of U. For example, for $n=1$, $U=\mathbb{C}^*$, and $A(z,t) = \mathrm{Diag}(z^{-2}, tz^{-2}, -tz^{-2})$, we find

three exponentials: $e^{-1/z}, e^{t/z}$ and $e^{-t/z}$. For $t \in \mathbb{R}^{>0}$, the singular directions of $(2t)z^{-1}$ are the same as the singular directions of $(t+1)z^{-1}$.

Let $(d_i(t))_{i\in\mathbb{N}}$ be the singular directions, and

$$\mathcal{D} = \{t \in U \mid \text{there exist } j, j' \in \mathbb{N} \text{ such that } d_j \not\equiv d_{j'} \text{ and } d_j(t) = d_{j'}(t)\}.$$

Lemma 1.12. *$\mathcal{D}$ is a closed subset of U with empty interior.*

Proof. Assume that there exist a nonempty polydisc $D \subset \mathcal{D}$ and two singular directions $d_j(t), d_{j'}(t)$ such that $d_j(t) = d_{j'}(t)$ on D. Then there exist a nonempty polydisc $D' \subset D$ and $q(t), q'(t) \in \mathcal{M}_{D'}$ that do not vanish on D' such that $q(t)/q'(t)$ has constant argument on D'. An analytic function with constant argument on a polydisc is constant. Hence, we deduce that $d_j(t) = d_{j'}(t)$ on a polydisc, which implies that $d_j(t) = d_{j'}(t)$ on U. Since the set of singular directions is finite modulo $2\pi\nu$ with $\nu \in \mathbb{N}^*$, $\mathcal{D}$ has empty interior. □

Thus, if we take a smaller nonempty polydisc U, we may assume the following:

- $\mathcal{D} = \varnothing$.
- The levels of $(*)$ are independent of t
- For all $t_0 \in U$, the singular directions of $(*)$ evaluated at t_0 are equal to the singular directions of the specialized system $\partial_z Y(z, t_0) = A(z, t_0)Y(z, t_0)$.

Let $\hat{H}(z,t)z^{L(t)}e(Q(z,t)) \in \mathrm{GL}_m((\mathbf{K}_U)^\wedge)$ be a fundamental solution to the parametrized linear differential system $(*)$ in the same form as in Proposition 1.3, where we consider $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$. Let $d(t)$ be a singular direction, and let $k_1 < \cdots < k_r$ be the levels of $(*)$. For t belonging to U, we define the parametrized Stokes matrix $\mathrm{St}_{d(t)}$ (we will just call it the Stokes matrix when no confusion is likely to arise) as $t \mapsto \mathrm{St}_{d(t)}$, where $\mathrm{St}_{d(t)}$ is the Stokes matrix of the specialized system defined just before Proposition 1.8.

Proposition 1.13. *Let $d(t)$ be continuous in t such that for all t_0 in U, $d(t_0)$ is not a singular direction of the unparametrized linear differential equation $\partial_z Y(z, t_0) = A(z, t_0)Y(z, t_0)$. We define $t \mapsto H^{d(t)}(z,t)e^{L(t)\log(z)}e^{Q(z,t)}$ as the solution* (1-2) *of the specialized system. Let $d_1(t), d_2(t)$ be two singular directions such that for all $t \in U$, $d_1(t) < d(t) < d_2(t)$ and $]d_1(t), d_2(t)[$ contains no singular directions. Then, there exists a map $U \to \mathbb{R}^{>0}$, $t \mapsto \varepsilon(t)$, which is not necessarily continuous, such that $H^{d(t)}(z,t)e^{L(t)\log(z)}e^{Q(z,t)}$ is meromorphic in (z,t) for*

$$z \in \Sigma(d_1(t) - \pi/2k_r, d_2(t) + \pi/2k_r) \text{ with } 0 < |z| < \varepsilon(t) \quad \text{and} \quad t \in U.$$

Notice that the facts that $\mathcal{D} = \varnothing$ and that the singular directions are continuous in t implies the existence of a continuous function $d(t)$ such that, for all t_0 in U, $d(t_0)$ is not a singular direction of the unparametrized linear differential equation $\partial_z Y(z, t_0) = A(z, t_0)Y(z, t_0)$.

Proof. We recall that we have assumed that for all $t_0 \in U$, the singular directions of $(*)$ evaluated at t_0 are equal to the singular directions of the specialized system $\partial_z Y(z, t_0) = A(z, t_0)Y(z, t_0)$. We have seen in Section 1C that, for t fixed, the asymptotic solution is a germ of meromorphic function on the sector

$$\Sigma(d_1(t) - \pi/2k_r, d_2(t) + \pi/2k_r).$$

We may replace $d(t)$ by any function, possibly discontinuous, such that for all $t \in U$, $d_1(t) < d(t) < d_2(t)$. Since the singular directions are continuous in t, we may assume that $d(t)$ is locally constant. Since for $z \neq 0$, $t \mapsto e^{L(t)\log(z)}e^{Q(z,t)} \in \mathcal{M}_U$, this is now a consequence of Proposition 1.10 and Lemma 1.14 below. □

Lemma 1.14. *We keep the same notation as in Definition 1.9 and Proposition 1.10. Let $\hat{h}(z, t)$ be one of the entries of $\hat{H}(z, t)$. Let $V \subset U$ be a nonempty polydisc, and let $d \in \mathbb{R}$ such that for all $t \in V$, d is not an unparametrized singular direction of $(*)$. Then there exists a map $U \to \mathbb{R}^{>0}$, $t \mapsto \varepsilon(t)$, which is not necessarily continuous, such that*

$$\mathcal{L}_{\kappa_r,d} \circ \cdots \circ \mathcal{L}_{\kappa_1,d} \circ \hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}(\hat{h})$$

is meromorphic in (z, t) for

$$z \in \Sigma(d - \pi/2k_r, d + \pi/2k_r) \text{ with } 0 < |z| < \varepsilon(t) \quad \text{and} \quad t \in V.$$

Moreover, for all $j \leq n$,

$$\mathcal{L}_{\kappa_r,d} \circ \cdots \circ \mathcal{L}_{\kappa_1,d} \circ \hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}(\partial_{t_j}\hat{h}) = \partial_{t_j}(\mathcal{L}_{\kappa_r,d} \circ \cdots \circ \mathcal{L}_{\kappa_1,d} \circ \hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}(\hat{h})).$$

Proof. We will proceed in two steps.

Step 1: We recall that $\hat{h}(z, t) \in \hat{K}_U[z^{1/\nu}]$ (where $\nu \in \mathbb{N}^*$ has been defined in Proposition 1.3) and (see Remark 1.7) all the z-coefficients are analytic on U. Because of Proposition 1.10, for t fixed, $\hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}(\hat{h})$ is a germ of a meromorphic function. Therefore, it belongs to $\mathcal{O}_U(\{z\})[z^{1/\nu}]$. Let h_1 be the analytic continuation defined in Proposition 1.10. In particular, for all $z \in \widetilde{\mathbb{C}}$ with $\arg(z) = d$, $t \mapsto h_1(z, t) \in \mathcal{M}_V$. The fact that we have a meromorphic function allows us to differentiate termwise, and for all $j \leq n$, $\partial_{t_j} h_1$ is equal to the analytic continuation of

$$\hat{\mathcal{B}}_{\kappa_r} \circ \cdots \circ \hat{\mathcal{B}}_{\kappa_1}(\partial_{t_j}\hat{h}).$$

Step 2: Let $h_2, \ldots, h_r$ be the successive Laplace transforms that were defined in Proposition 1.10. Let $t_0 \in V$, let W_{t_0} be a compact neighborhood of t_0 in V, let $i \leq r$, and assume that for $z \in \widetilde{\mathbb{C}}$ with $\arg(z) = d$, $t \mapsto h_i(z, t)$ is meromorphic on W_{t_0}. It is sufficient to prove that, for all $z \in \widetilde{\mathbb{C}}$ with $\arg(z) \in \,]d - \pi/2\kappa_i, d + \pi/2\kappa_i[$ and $|z|$ sufficiently small, $t \mapsto h_{i+1}(z, t)$ is meromorphic on W_{t_0}, and for all $j \leq n$,

$$\mathcal{L}_{\kappa_i,d}(\partial_{t_j} h_i) = \partial_{t_j}(\mathcal{L}_{\kappa_i,d}(h_i)) = \partial_{t_j} h_{i+1}.$$

The function $\mathscr{L}_{\kappa_i,d}(h_i)$ is an integral of a meromorphic function depending analytically upon parameters, and we just have to prove that it is possible to find a function f such that, for all $t \in W_{t_0}$, $|h_i(u,t)| < |f(u)|$ and for $\arg(z) \in]d-\pi/2\kappa_i, d+\pi/2\kappa_i[$, $|z|$ sufficiently small, $\mathscr{L}_{\kappa_i,d}(|f|)(z) < \infty$. From Proposition 1.10, we obtain the existence of $A(t), B(t) > 0$ such that for $\arg(u) = d$, $|h_i(u,t)| \leq A(t)e^{B(t)|u|^{\kappa_i}}$. Since $h_i(u,t)$ is meromorphic, we may assume that $A(t)$ and $B(t)$ are continuous on W_{t_0}. The functions $A(t)$ and $B(t)$ admit a maximum A and B on the compact set W_{t_0}. Finally, for $\arg(z) \in]d-\pi/2\kappa_i, d+\pi/2\kappa_i[$ and $|z|$ sufficiently small,

$$\begin{aligned}|\mathscr{L}_{\kappa_i,d}h_i| &= \left|\int_0^{\infty e^{id}} h_i(u,t)e^{-(u/z)^{\kappa_i}}\, d\left(\left(\frac{u}{z}\right)^{\kappa_i}\right)\right| \\ &\leq \int_0^{\infty} Ae^{B|u|^{\kappa_i}}|e^{-(u/z)^{\kappa_i}}|d\left(\left(\frac{u}{z}\right)^{\kappa_i}\right) < \infty. \qquad \square\end{aligned}$$

2. Parametrized differential Galois theory

In this section we are interested in parametrized differential Galois theory: this is a generalization of differential Galois theory for parametrized linear differential equations. In Section 2A, we review the parametrized differential Galois theory developed in [CS]. In Section 2B, we prove that some of the results of Section 2A stay valid without the assumption that the field of constants is differentially closed. This will help us in Section 2C to prove that the local analytic parametrized differential Galois group descends to a smaller field, whose field of constants is not differentially closed. In Section 2D, we explain the main result of the paper: we show an analogue of the density theorem of Ramis in the parametrized case. In Section 2E, we give a similar result for the global parametrized differential Galois group. We end by giving various examples of computation of parametrized differential Galois groups using the parametrized density theorem.

2A. ***Basic facts.*** We recall some facts from [CS] about Galois theory of parametrized linear differential equations. Classical Galois theory of unparametrized linear differential equation is presented in some books, such as [van der Put and Singer 2003; Magid 1994].

Let K be a differential field of characteristic 0 with $n+1$ commuting derivations $\partial_0, \dots, \partial_n$. We want to study differential equations of the form $\partial_0 Y = AY$, with $A \in \mathrm{M}_m(K)$. Let C_K be the field of constants with respect to ∂_0. Since all the derivations commute with ∂_0, $(C_K, \partial_1, \dots, \partial_n)$ is a differential field. By abuse of notation, we will sometimes start from a $(\partial_1, \dots, \partial_n)$-differential field C_K and build a $(\partial_0, \dots, \partial_n)$-differential field extension K of C_K, such that C_K is the field of constants with respect to ∂_0.

Example 2.1. If $K = \hat{K}_U$, then $\partial_0 = \partial_z$, $\{\partial_1, \dots, \partial_n\} = \Delta_t$, and $C_K = \mathscr{M}_U$.

A parametrized Picard–Vessiot extension for the parametrized linear differential equation $\partial_0 Y = AY$ on K is a $(\partial_0, \dots, \partial_n)$-differential field extension $\widetilde{K}|K$ with the following properties:

- There exists a fundamental solution for $\partial_0 Y = AY$ in $\widetilde{K}$, i.e., an invertible matrix $U = (u_{i,j})$, with entries in $\widetilde{K}$, such that $\partial_0 U = AU$.
- $\widetilde{K} = K\langle u_{i,j}\rangle_{\partial_0,\dots,\partial n}$, i.e., $\widetilde{K}$ is the $(\partial_0, \dots, \partial_n)$-differential field generated by K and the $u_{i,j}$.
- The field of constants of $\widetilde{K}$ with respect to ∂_0 is C_K.

Let L be a $(\partial_1, \dots, \partial_n)$-field of characteristic 0 with commuting derivations. The $(\partial_1, \dots, \partial_n)$-differential ring $L\{y_1, \dots, y_k\}_{\partial_1,\dots,\partial_n}$ of differential polynomials in k indeterminates over L is the usual polynomial ring in the infinite set of variables

$$\left\{\partial_1^{\nu_1} \dots \partial_n^{\nu_n} y_j\right\}_{j\le k}^{\nu_i\in\mathbb{N}},$$

and with derivations extending those in $\{\partial_1, \dots, \partial_n\}$ on L, defined by

$$\partial_i(\partial_1^{\nu_1} \dots \partial_n^{\nu_n} y_j) = \partial_1^{\nu_1} \dots \partial_i^{\nu_i+1} \dots \partial_n^{\nu_n} y_j.$$

Definition 2.2 [Cassidy and Singer 2007, Definition 3.2]. We say that the field $(C_K, \partial_1, \dots, \partial_n)$ is differentially closed if it has the following property: for any $k, l \in \mathbb{N}$ and for all $P_1, \dots, P_k \in C_K\{y_1, \dots, y_l\}_{\partial_1,\dots,\partial_n}$, the system

$$\begin{cases} P_1(\alpha_1, \dots, \alpha_l) = 0 \\ \quad\vdots \\ P_{k-1}(\alpha_1, \dots, \alpha_l) = 0 \\ P_k(\alpha_1, \dots, \alpha_l) \neq 0, \end{cases}$$

has a solution in C_K as soon as it has a solution in a $(\partial_1, \dots, \partial_n)$-differential field containing C_K.

For simplicity of notation, we will say that C_K is differentially closed rather than that $(C_K, \partial_1, \dots, \partial_n)$ is differentially closed. Note that there exists a differentially closed extension of C_K; see [CS, Section 9.1]. By definition, a differentially closed field is algebraically closed.

Proposition 2.3 [CS, Theorem 9.5]. *Assume that C_K is differentially closed. Then the parametrized Picard–Vessiot extension for $\partial_0 Y = AY$ exists and is unique up to $(\partial_0, \dots, \partial_n)$-differential isomorphism.*

Until the end of the Section 2A, we assume that C_K is differentially closed.

Consider $\partial_0 Y = AY$ with $A \in \mathrm{M}_m(K)$, and let $\widetilde{K}|K$ be a parametrized Picard–Vessiot extension. The parametrized differential Galois group $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$ is the group of field automorphisms of $\widetilde{K}$ which induce the identity on K and commute with all the derivations. This latter is independent of the choice of the parametrized Picard–Vessiot extension, since all the parametrized Picard–Vessiot extensions are $(\partial_0, \dots, \partial_n)$-differentially isomorphic. In the unparametrized case, the differential Galois group is an algebraic subgroup of $\mathrm{GL}_m(C_K)$. In the parametrized case, we find a linear differential algebraic subgroup:

Definition 2.4. Let us consider m^2 indeterminates $(X_{i,j})_{i,j\le m}$. We say that a subgroup G of $\mathrm{GL}_m(C_K)$ is a linear differential algebraic group if there exist $P_1, \dots, P_k \in C_K\{X_{i,j}\}_{\partial_1,\dots,\partial_n}$ such that for $A = (a_{i,j}) \in \mathrm{GL}_m(C_K)$,

$$A \in G \iff P_1(a_{i,j}) = \cdots = P_k(a_{i,j}) = 0.$$

Let U be a fundamental solution of $\partial_0 Y = AY$. One proves directly that the map

$$\begin{aligned}\rho_U : \mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K) &\longrightarrow \mathrm{GL}_m(C_K)\\ \varphi &\longmapsto U^{-1}\varphi(U),\end{aligned}$$

is an injective group morphism. A fundamental fact is that

$$\operatorname{Im}\rho_U = \{U^{-1}\varphi(U) \mid \varphi \in \mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)\}$$

is a linear differential algebraic subgroup of $\mathrm{GL}_m(C_K)$ (see Theorem 9.5 in [CS]). If we take a different fundamental solution in $\widetilde{K}$, we obtain a conjugate linear differential algebraic subgroup of $\mathrm{GL}_m(C_K)$. We will identify $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$ with a linear differential algebraic subgroup of $\mathrm{GL}_m(C_K)$ for a chosen fundamental solution. We put a topology on $\mathrm{GL}_m(C_K)$, called the Kolchin topology, for which the closed sets are defined as the zero loci of finite sets of differential polynomials with coefficients in C_K.

Example 2.5 [CS, Example 3.1]. Let $n=1$, let (C_K, ∂_t) be a differentially closed ∂_t-field that contains $(\mathbb{C}(t), \partial_t)$, and let us consider $K = C_K(z)$, the (∂_z, ∂_t)-differential field of rational functions in the indeterminate z with coefficients in C_K, where z is a ∂_t-constant with $\partial_z z = 1$, C_K is the field of constants with respect to ∂_z, and ∂_z commutes with ∂_t. Let us consider the parametrized differential equation

$$\partial_z Y(z,t) = \frac{t}{z} Y(z,t).$$

The fundamental solution is (z^t), and $K(z^t, \log)$ is a parametrized Picard–Vessiot extension (see Section 1A for the notation). Here we have added log because we want the extension to be closed under the derivations ∂_z and ∂_t. Using the fact that the Galois group commutes with ∂_z and ∂_t, we find that the Galois group is given by

$$\{f \in C_K \mid f \neq 0 \text{ and } f\partial_t^2 f - (\partial_t f)^2 = 0\}.$$

We can see that if we take $C_K = \mathbb{C}(t)$ or $C_K = \mathcal{M}_{\mathbb{C}}$ (see page 90), which are not differentially closed, then we find two different groups of differential automorphisms:

$$\{f \in \mathbb{C}(t) \mid f \neq 0 \text{ and } f\partial_t^2 f - (\partial_t f)^2 = 0\} = \mathbb{C}^*$$

and

$$\{f \in \mathcal{M}_{\mathbb{C}} \mid f \neq 0 \text{ and } f\partial_t^2 f - (\partial_t f)^2 = 0\} = \{ce^{bt} \mid b \in \mathbb{C}, c \in \mathbb{C}^*\},$$

which shows the importance of considering a Galois group defined over a differentially closed field. See Example 2.26 for the resolution of this ambiguity using the parametrized density theorem.

There is a Galois correspondence theorem for parametrized differential Galois theory; see Theorem 9.5 in [CS]. For a subgroup G of $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$, let

$$\widetilde{K}^G = \{a \in \widetilde{K} \mid \sigma(a) = a \text{ for all } \sigma \in G\}.$$

Then the theorem says that the Kolchin-closed subgroups of $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$ are in bijection with the $(\partial_0, \dots, \partial_n)$-differential subfields of $\widetilde{K}$ containing K via the map

$$G \mapsto \widetilde{K}^G.$$

The inverse map is given by

$$M \mapsto \mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|M),$$

where $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|M)$ denotes the set of elements of $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$ inducing the identity map on M. In particular, we have the following corollary:

Corollary 2.6. *Let G be an arbitrary subgroup of* $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$. *Then $\widetilde{K}^G = K$ if and only if G is dense for the Kolchin topology in* $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$.

Let $L|M|K$ be $(\partial_1, \dots, \partial_n)$-differential field extensions. Notice that we do not exclude $L = M = K$. All the definitions that we give before the next proposition come from [Hardouin and Singer 2008, §6.2.3].

We remark that $P(a_1, \dots, a_n)$ is well-defined for $P \in M\{X_1, \dots, X_k\}_{\partial_1,\dots,\partial_n}$ and $a_1, \dots, a_k \in L$. Then we may define the $(\partial_1, \dots, \partial_n)$-differential transcendence degree of L over M as the maximum number of elements $a_1, \dots, a_k$ of L such that

$$P(a_1, \dots, a_k) \neq 0,$$

for all nonzero $(\partial_1, \dots, \partial_n)$-differential polynomials P with coefficients in M. The $(\partial_1, \dots, \partial_n)$-differential transcendence degree of an integral domain over another integral domain is defined to be the $(\partial_1, \dots, \partial_n)$-differential transcendence degree of the fraction field of the first one over the fraction field of the second one.

Let us consider m^2 indeterminates $(X_{i,j})_{i,j\leq m}$. Let (p) be a prime $(\partial_1, \dots, \partial_n)$-differential ideal of $C_K\{X_{i,j}\}_{\partial_1,\dots,\partial_n}$, i.e., a prime ideal stable under the derivations $\partial_1, \dots, \partial_n$. The $(\partial_1, \dots, \partial_n)$-dimension of (p) over C_K is defined to be the $(\partial_1, \dots, \partial_n)$-differential transcendence degree of the quotient ring

$$C_K\{X_{i,j}\}_{\partial_1,\dots,\partial_n}/(p)$$

over C_K.

Let (r) be a radical $(\partial_1, \dots, \partial_n)$-differential ideal of $C_K\{X_{i,j}\}_{\partial_1,\dots,\partial_n}$, i.e., a radical ideal stable under the derivations $\partial_1, \dots, \partial_n$. Let $(p_1), \dots, (p_\nu)$ with $\nu \in \mathbb{N}^*$ be the prime $(\partial_1, \dots, \partial_n)$-differential ideals such that $(r) = \bigcap_{k\leq\nu}(p_k)$. The $(\partial_1, \dots, \partial_n)$-dimension of (r) over C_K is defined to be the maximum in k of the $(\partial_1, \dots, \partial_n)$-dimension of (p_k) over C_K.

Assume that $M \subset \widetilde{K}$. Let (q) be the radical $(\partial_1, \dots, \partial_n)$-differential ideal of $C_K\{X_{i,j}\}_{\partial_1,\dots,\partial_n}$ that defines $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|M)$ (see the proof of Proposition 9.10 in [CS]). We define the $(\partial_1, \dots, \partial_n)$-differential dimension of $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|M)$ over C_K as the $(\partial_1, \dots, \partial_n)$-dimension of (q) over C_K.

Proposition 2.7 [Hardouin and Singer 2008, Proposition 6.26]. *The $(\partial_1, \dots, \partial_n)$-differential transcendence degree of $\widetilde{K}$ over M is equal to the $(\partial_1, \dots, \partial_n)$-differential dimension of $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|M)$ over C_K.*

Example 2.5 revisited. Let us keep the same notation as in Example 2.5. The parametrized Picard–Vessiot extension is $K(z^t, \log)$ and the Galois group is

$$\{f \in C_K \mid f \neq 0 \text{ and } f\partial_t^2 f - (\partial_t f)^2 = 0\}.$$

We may directly check that the ∂_t-differential dimension of the Galois group is 0, and therefore z^t satisfies a ∂_t-differential polynomial equation with coefficients in C_K.

2B. ***Parametrized differential Galois theory for a nondifferentially closed field of constants.*** Let K be a differential field of characteristic 0 with $n+1$ commuting derivations $\partial_0, \dots, \partial_n$. Let C_K be the field of constants with respect to ∂_0. Note that we do not assume C_K to be differentially closed. Consider $\partial_0 Y = AY$, with $A \in \mathrm{M}_m(K)$, and assume the existence of $\widetilde{K}|K$, a parametrized Picard–Vessiot extension for $\partial_0 Y = AY$ (see Section 2A). This means in particular that the field of constants of $\widetilde{K}$ with respect to ∂_0 is C_K. Let $F = (F_{i,j}) \in \mathrm{GL}_m(\widetilde{K})$ be a fundamental solution such that $\widetilde{K} = K\langle F_{i,j}\rangle_{\partial_0,\dots,\partial_n}$ (see Section 2A for the notation). Let $\mathrm{Aut}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$ be the group of $(\partial_0, \dots, \partial_n)$-differential field automorphisms of $\widetilde{K}$ keeping K invariant.

Remark 2.8. We avoid here the notation $\mathrm{Gal}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$, because we have no theorem that guarantees the uniqueness of the parametrized Picard–Vessiot extension

$\widetilde{K}|K$, since C_K is not differentially closed. However, we will call it the parametrized differential Galois group, or Galois group, if no confusion is likely to arise.

We extend Definition 2.4 for the field C_K. Let us consider m^2 indeterminates $(X_{i,j})_{i,j\leq m}$. We say that a subgroup G of $\mathrm{GL}_m(C_K)$ is a linear differential algebraic group if there exist $P_1, \dots, P_k \in C_K\{X_{i,j}\}_{\partial_1,\dots,\partial_n}$ such that for $A = (a_{i,j}) \in \mathrm{GL}_m(C_K)$,

$$A \in G \iff P_1(a_{i,j}) = \cdots = P_k(a_{i,j}) = 0.$$

The goal of the subsection is to prove:

Proposition 2.9. (1) *Let us consider the injective group morphism*

$$\begin{aligned} \rho_F : \mathrm{Aut}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K) &\longrightarrow \mathrm{GL}_m(C_K) \\ \varphi &\longmapsto F^{-1}\varphi(F). \end{aligned}$$

Then

$$\mathrm{Im}\,\rho_F = \{F^{-1}\varphi(F) \mid \varphi \in \mathrm{Aut}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)\}$$

is a linear differential algebraic subgroup of $\mathrm{GL}_m(C_K)$. *We will identify* $\mathrm{Aut}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$ *with a linear differential algebraic subgroup of* $\mathrm{GL}_m(C_K)$ *for a chosen fundamental solution. The image is independent of this choice, up to conjugacy by an element of* $\mathrm{GL}_m(C_K)$.

(2) *Let G be a subgroup of* $\mathrm{Aut}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$. *If* $\widetilde{K}^G = K$, *then G is dense in* $\mathrm{Aut}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$ *for the Kolchin topology.*

Remark that, contrary to Corollary 2.6, the converse of (2) is false when C_K is not differentially closed. See [CS, Example 3.1]. Before showing the proposition, we point out two facts we will use in the proof. Let $L|K$ be a $(\partial_0, \dots, \partial_n)$-differential field extension and $a_1, \dots, a_k \in L$.

- As in the case where C_K is differentially closed (see Section 2A), $P(a_1, \dots, a_k)$ is well-defined for $P \in K\{X_1, \dots, X_k\}_{\partial_1,\dots,\partial_n}$.
- The set $\{P(a_1, \dots, a_k) \mid P \in K\{X_1, \dots, X_k\}_{\partial_1,\dots,\partial_n}\}$ is a $(\partial_0, \dots, \partial_n)$-differential field extension we will denote by $L\{a_1, \dots, a_k\}_{\partial_1,\dots,\partial_n}|L$.

Proof of Proposition 2.9.

Part (1): We follow here the proof of Proposition 9.10 in [CS]. We consider the differential polynomial ring

$$R = K\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n},$$

and endow it with the ∂_0-differential structure defined by $\partial_0(X_{i,j}) = A(X_{i,j})$. Let us consider

$$S = K\{F_{i,j}, 1/\det(F_{i,j})\}_{\partial_0,\dots,\partial_n},$$

the $(\partial_0, \dots, \partial_n)$-differential subring of $\widetilde{K}$ generated by the $F_{i,j}$ and $1/\det(F_{i,j})$ over K. It is an integral domain. Let q be the obvious prime $(\partial_0, \dots, \partial_n)$-differential ideal such that $R/q \simeq S$. Let $Z_{i,j}$ be the image of $X_{i,j}$ in $S \subset \widetilde{K}$, so that $(Z_{i,j})$ is a fundamental solution for $\partial_0 Y = AY$ in S. Consider the following rings:

$$\begin{array}{ccc} \widetilde{K}\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} & = & \widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n} \\ \cup & & \cup \\ K\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} & & C_K\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}, \end{array}$$

where the indeterminates $Y_{i,j}$ are defined by $(X_{i,j}) = (Z_{i,j})(Y_{i,j})$. We remark that $\partial_0(Y_{i,j}) = 0$. Since we consider fields that are of characteristic 0, the differential ideal

$$\begin{aligned} q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n} \subset \widetilde{K}\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} \\ = \widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n} \end{aligned}$$

is a radical $(\partial_0, \dots, \partial_n)$-differential ideal (see Corollary A.17 in [van der Put and Singer 2003]). The next lemma is an adaptation of Lemma 9.8 in [CS] without the assumption that the field of constants is differentially closed.

Lemma 2.10. *The* $(\partial_0, \dots, \partial_n)$*-ideal* $q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$ *is generated by*

$$I = q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n} \cap C_K\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}.$$

Proof. Let $(e_i)_{i\in B}$ be a basis of $C_K\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$ over C_K. Let

$$f = \sum_{i=1}^{n} m_i e_i \in q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n},$$

with $m_i \in \widetilde{K}$. By induction on n we will show that $f \in I$. If $n = 0$ or 1 there is nothing to prove. We assume that $n > 1$. We can suppose that $m_1 = 1$ and $m_2 \notin C_K$. Then, since the field of constants of $\widetilde{K}$ with respect to ∂_z is C_K,

$$\partial_0(f) = \sum_{i=2}^{n} \partial_0(m_i)e_i \neq 0 \quad \text{and} \quad f \in q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}.$$

Then, by induction, $\partial_0(f) \in I$. By the same argument,

$$\partial_0(m_2^{-1} f) \in I.$$

Then $\partial_0(m_2^{-1})f = \partial_0(m_2^{-1} f) - m_2^{-1}\partial_0 f \in I$. Since $\partial_0(m_2^{-1}) \neq 0$, we obtain that $f \in I$. □

By Lemma 2.10, $q\widetilde{K}\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n}$ is generated by

$$I = q\widetilde{K}\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} \cap C_K\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}.$$

Clearly, I is a $(\partial_1, \dots, \partial_n)$-radical ideal of $C_K\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$. Let $C = (C_{i,j}) \in \mathrm{GL}_m(C_K)$. The following statements are equivalent:

(1) $(C_{i,j}) \in \mathrm{Aut}_{\partial_0}^{\partial_1,\dots,\partial_n}(\widetilde{K}|K)$.

(2) The map $K\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} \to K\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n}$ defined by $(X_{i,j}) \mapsto (X_{i,j})(C_{i,j}) := (\sum_{k=1}^m X_{i,k}C_{k,j})$ leaves q invariant.

(3) The map $K\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} \to \widetilde{K}$ defined by $(X_{i,j}) \mapsto (Z_{i,j})(C_{i,j})$ sends q to 0.

(4) The map $\widetilde{K}\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} \to \widetilde{K}$ defined by $(X_{i,j}) \mapsto (Z_{i,j})(C_{i,j})$ sends $q\widetilde{K}\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n} = q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$ to 0.

(5) The map $\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n} \to \widetilde{K}$ defined by $(Y_{i,j}) \mapsto (C_{i,j})$ sends $q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$ to 0.

The theorem is now a consequence of the fact that $q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$ is generated by I, a $(\partial_1, \dots, \partial_n)$-radical ideal of $C_K\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$.

Part (2): We follow the proof of Proposition 9.10 in [CS], and use the same notation as before. By construction, the ideal I of Lemma 2.10 above is the differential ideal that defines the Galois group. Assume that the Kolchin closure of G is not the whole Galois group. Then there exists $P \in C_K\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}$ such that $P \notin I$ and $P(g) = 0$ for all $g \in G$. Lemma 2.10 implies that

$$P \notin J = q\widetilde{K}\{Y_{i,j}, 1/\det(Y_{i,j})\}_{\partial_1,\dots,\partial_n}.$$

Let T consist of all $Q \in \widetilde{K}\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n}$ such that $Q \notin J$ and

$$Q((Z_{i,j})(g_{i,j})) = 0 \quad \text{for all } g = (g_{i,j}) \in G.$$

Since $P \in T$, $T \neq \{0\}$. An element $Q \in T$ can be written as

$$Q = f_1 Q_1 + \cdots + f_\nu Q_\nu,$$

where $f_i \in \widetilde{K}$ and $Q_i \in K\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n}$. Let $Q = f_1Q_1 + \cdots + f_\nu Q_\nu \in T$ such that:

- $f_1 = 1$,
- all the f_i are nonzero,
- ν is minimal.

For all $g \in G$, let $Q^g = f_1^g Q_1 + \cdots + f_\nu^g Q_\nu \in T$. Let $g \in G$. Since $Q - Q^g$ is shorter than Q, and satisfies $(Q - Q^g)((Z_{i,j})(g_{i,j})) = 0$, we have $Q - Q^g \in J$. If $Q - Q^g \neq 0$, there exists $l \in \widetilde{K}$ such that $Q - l(Q - Q^g)$ is shorter than Q. Since $Q - l(Q - Q^g) \in T$, this is not possible unless $Q - Q^g = 0$. Therefore, $Q = Q^g$, for all $g \in G$, and so $Q \in K\{X_{i,j}, 1/\det(X_{i,j})\}_{\partial_1,\dots,\partial_n}$. Since $Q(Z_{i,j}) = 0$, we have $Q \in J$. This completes the proof of Proposition 2.9. □

2C. ***A result of descent for the local analytic parametrized differential Galois group.*** We keep the notations of Section 1. Consider the equation $(*)$ with $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$, where U is a nonempty polydisc in $\mathbb{C}^n$, and $\mathbb{O}_U(\{z\})$ has been defined on page 90.

Remark 2.11. Note that $\mathbb{O}_U(\{z\})$ is a ring but not a field in general. For example, if $n=1$, $(z-t)^{-1} \notin \mathbb{O}_U(\{z\})$. However, we have $(z-t)^{-1} \in \mathbb{O}_{\mathbb{C}^*}(\{z\})$. More generally let $\alpha(z,t) \in \mathbb{O}_U(\{z\})$. For $t \in U$, let be $R(t)$ minimal such that $|\alpha(z,t)| \neq 0$ for $0<|z|<R(t)$. There exist a nonempty polydisc U' and $\varepsilon > 0$ with $R(t) > \varepsilon$ on U'. In particular, we have $\alpha(z,t)^{-1} \in \mathbb{O}_{U'}(\{z\})$.

Since $\mathbb{O}_U(\{z\}) \subset \hat{K}_U$, which is a field, $\mathbb{O}_U(\{z\})$ is an integral domain, and we can define K_U as the fraction field of $\mathbb{O}_U(\{z\})$. We have

$$\{a \in K_U \mid \partial_z a = 0\} = \{a \in \hat{K}_U \mid \partial_z a = 0\} = \mathcal{M}_U.$$

Let

$$F(z,t) = (F_{i,j}) = \hat{H}(z,t) z^{L(t)} e(Q(z,t)) \in \mathrm{GL}_m((\mathbf{K}_U)^{\wedge}) \quad \text{(see Section 1A)}$$

be the fundamental solution given in Proposition 1.3. Let us denote

$$K_U \langle F_{i,j} \rangle_{\partial_z, \Delta_t} = (K_U)^{\sim},$$

which is a (∂_z, Δ_t)-differential subfield of $(\mathbf{K}_U)^{\wedge}$. We have seen in Section 1A that $(\mathbf{K}_U)^{\wedge}$ has field of constants with respect to ∂_z equal to $\mathcal{M}_U$. Then we deduce that $(K_U)^{\sim}|K_U$ is a parametrized Picard–Vessiot extension. Therefore, the results of Section 2B may be applied here; and we can define a parametrized differential Galois group $\mathrm{Aut}^{\Delta_t}_{\partial_z}((K_U)^{\sim}|K_U)$, which will be identified with a linear differential algebraic subgroup of $\mathrm{GL}_m(\mathcal{M}_U)$. We want to prove now that it is the "same" as the one of Section 2A.

Let C be a (Δ_t)-differentially closed field that contains $\mathcal{M}_U$. Let us define $C[\![z]\!][z^{-1}]$, the (∂_z, Δ_t)-differential field, where z is a (Δ_t)-constant with $\partial_z z = 1$, C is the field of constants with respect to ∂_z, and ∂_z commutes with all the derivations. We define the ring $K_U \otimes_{\mathcal{M}_U} C$ with the differential structure given by

$$\partial(a \otimes_{\mathcal{M}_U} c) = \partial a \otimes_{\mathcal{M}_U} c + a \otimes_{\mathcal{M}_U} \partial c \quad \text{for all } a \in K_U, c \in C, \partial \in \{\partial_z, \Delta_t\}.$$

This (∂_z, Δ_t)-differential ring can be naturally embedded into $C[\![z]\!][z^{-1}]$, which implies that it is an integral domain. Therefore we may define $\mathcal{K}_{C,U}$, the field of fractions of $K_U \otimes_{\mathcal{M}_U} C$. We see now $\mathcal{K}_{C,U}$ (resp. $K_U \otimes_{\mathcal{M}_U} C$) as a subfield (resp. subring) of $C[\![z]\!][z^{-1}]$.

Proposition 2.12. *Let us keep the same notation. Consider the equation* $\partial_z Y(z,t) = A(z,t)Y(z,t)$ *with* $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$. *The extension field*

$$\mathcal{K}_{C,U} \langle F_{i,j} \rangle_{\partial_z, \Delta_t} | \mathcal{K}_{C,U} = (\mathcal{K}_{C,U})^{\sim} | \mathcal{K}_{C,U}$$

is a parametrized Picard–Vessiot extension for $\partial_z Y(z,t) = A(z,t)Y(z,t)$*. Moreover, there exist* $P_1, \dots, P_k \in \mathcal{M}_U\{X_{i,j}\}_{\Delta_t}$ *such that the image of the representation of* $\mathrm{Gal}_{\partial_z}^{\Delta_t}((\mathcal{K}_{C,U})^{\sim}|\mathcal{K}_{C,U})$ (*resp.* $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$) *associated to* $F(z,t)$ *is the set of* C*-rational points* (*resp.* $\mathcal{M}_U$*-rational points*) *of the linear differential algebraic subgroup of* $\mathrm{GL}_m(C)$ (*resp.* $\mathrm{GL}_m(\mathcal{M}_U)$) *defined by* $P_1, \dots, P_k$*. More explicitly,*

$$\{F^{-1}\varphi(F) \mid \varphi \in \mathrm{Gal}_{\partial_z}^{\Delta_t}((\mathcal{K}_{C,U})^{\sim}|\mathcal{K}_{C,U})\} \\ = \{A = (a_{i,j}) \in \mathrm{GL}_m(C) \mid P_1(a_{i,j}) = \cdots = P_k(a_{i,j}) = 0\}$$

and

$$\{F^{-1}\varphi(F) \mid \varphi \in \mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)\} \\ = \{A = (a_{i,j}) \in \mathrm{GL}_m(\mathcal{M}_U) \mid P_1(a_{i,j}) = \cdots = P_k(a_{i,j}) = 0\}.$$

Proof. We follow the proof of [Mitschi and Singer 2012, Proposition 3.3]. Let (d_k) be an $\mathcal{M}_U$-basis of C. Let us prove that the d_k are linearly independent over $(K_U)^{\sim}$. Write $\sum_{k\leq\kappa} d_k P_k = 0$ with $0 \neq P_k \in (K_U)^{\sim}$, $\kappa \geq 2$ minimal and $P_\kappa = 1$. We have $\sum_{k\leq\kappa-1} d_k \partial_z P_k = 0$. If $\kappa = 2$, then $\partial_z P_1 = 0$. If $\kappa > 2$, the minimality of κ implies that $\partial_z P_k = 0$ for all k. Since $(K_U)^{\sim}|K_U$ is a parametrized Picard–Vessiot extension, $P_k \in \mathcal{M}_U$ for all k, and the d_k are linearly independent over $(K_U)^{\sim}$.

Now, we prove that $\mathcal{K}_{C,U}\langle F_{i,j}\rangle_{\partial_z,\Delta_t}|\mathcal{K}_{C,U}$ is a parametrized Picard–Vessiot extension for $\partial_z Y(z,t) = A(z,t)Y(z,t)$. Let $\alpha \in \mathcal{K}_{C,U}\langle F_{i,j}\rangle_{\partial_z,\Delta_t}$ with $\partial_z\alpha = 0$. We may assume that $\alpha = \sum d_k P_k$, where $P_k \in (K_U)^{\sim}$. We have $\partial_z\alpha = \sum d_k \partial_z P_k = 0$. Since the d_k are linearly independent over $(K_U)^{\sim}$, we find $\partial_z P_k = 0$. Hence, $P_k \in \mathcal{M}_U$, because $(K_U)^{\sim}|K_U$ is a parametrized Picard–Vessiot extension. Therefore, $\alpha \in C$ and $\mathcal{K}_{C,U}\langle F_{i,j}\rangle_{\partial_z,\Delta_t}|\mathcal{K}_{C,U}$ is a parametrized Picard–Vessiot extension for $\partial_z Y(z,t) = A(z,t)Y(z,t)$.

Let $Y_{i,j}$ be a set of m^2 indeterminates and let I_0, I_1 be (∂_z, Δ_t)-differential ideals such that

$$R_0 = K_U\{F_{i,j}\}_{\partial_z,\Delta_t} = K_U\{Y_{i,j}\}_{\partial_z,\Delta_t}/I_0,$$
$$R_1 = \mathcal{K}_{C,U}\{F_{i,j}\}_{\partial_z,\Delta_t} = \mathcal{K}_{C,U}\{Y_{i,j}\}_{\partial_z,\Delta_t}/I_1.$$

The group $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$ (resp. $\mathrm{Gal}_{\partial_z}^{\Delta_t}((\mathcal{K}_{C,U})^{\sim}|\mathcal{K}_{C,U})$) is the set of $B \in \mathrm{GL}_m(\mathcal{M}_U)$ (resp. $B \in \mathrm{GL}_m(C)$) such that $(F_{i,j})B$ is again a zero of I_0 (resp. I_1). We just have to prove that $I_1 = CI_0$. The inclusion $CI_0 \subset I_1$ is clear. Let us prove the other inclusion. Let $P \in I_1$. Without loss of generality, we may assume that $P \in (K_U \otimes_{\mathcal{M}_U} C)[Y_{i,j}]$. Let us write $P = \sum d_k P_k$, where $P_k \in K_U[Y_{i,j}]$. One finds that

$$P(F_{i,j}) = \sum d_k P_k(F_{i,j}) = 0.$$

Since the d_k are linearly independent over $(K_U)^{\sim}$, one finds that $P_k(F_{i,j}) = 0$, and therefore $I_1 = C I_0$. $\square$

2D. ***An analogue of the density theorem in the parametrized case.*** Let us consider the equation $(*)$, with $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$, where U is a nonempty polydisc in $\mathbb{C}^n$. We want to find topological generators for $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$ for the Kolchin topology.

We now define the parametrized monodromy. The notion of monodromy in the unparametrized case is well explained in [van der Put and Singer 2003]. For more details about parametrized monodromy, see [Cassidy and Singer 2007; Mitschi and Singer 2012; 2013; Sibuya 1990].

Definition 2.13. The notations are introduced in Section 1A. We define $\hat{m}$, the formal parametrized monodromy, as follows:

- $\hat{m}(\hat{H}(z,t)) = \hat{H}(z,t)$ for all $\hat{H}(z,t) \in \hat{K}_U$.
- $\hat{m}(z^{a(t)}) = e^{2i\pi a(t)} z^{a(t)}$ for all $a(t) \in \mathcal{M}_U$.
- $\hat{m}(\log) = 2i\pi + \log$.
- For all $q(z,t) = \sum a_n z^{-n} \in \mathbf{E}_U = \bigcup_{\nu \in \mathbb{Q}^{>0}} z^{-1/\nu} \mathcal{M}_U[z^{-1/\nu}]$, we define
$$\hat{m}(e(q(z,t))) = e\Big(\sum a_n e^{-2i\pi n} z^{-n}\Big).$$

From the construction of $\hat{K}_U[\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U} (e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}]$, it is easy to check that $\hat{m}$ induces a well defined (∂_z, Δ_t)-differential ring automorphism of $\hat{K}_U[\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U} (e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}]$, and then it can be extended as a (∂_z, Δ_t)-differential field automorphism of $(\mathbf{K}_U)^{\wedge}$ keeping K_U invariant. Since $(K_U)^{\sim} \subset (\mathbf{K}_U)^{\wedge}$, and since $(K_U)^{\sim}$ is stable under $\hat{m}$, $\hat{m}$ induces an element of $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$.

Remark 2.14. In the regular singular case with one singularity at 0, the definition of formal parametrized monodromy restricts to the definition given in [Mitschi and Singer 2012].

We now introduce the parametrized exponential torus, which is a subgroup of $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$ consisting of elements that act on the $e(q(z,t))$ with $q(z,t) \in \mathbf{E}_U$.

Definition 2.15. Let α be a character of $\mathbf{E}_U$. We define τ_α as follows:

- τ_α is the identity on $\hat{K}_{F,U}$.
- $\tau_\alpha(e(q(z,t))) = \alpha(q(z,t)) e(q(z,t))$ for all $q(z,t) \in \mathbf{E}_U$.

From the construction of $\hat{K}_U[\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U} (e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}]$, it is easy to check that τ_α induces a well defined (∂_z, Δ_t)-differential ring automorphism of $\hat{K}_U[\log, (z^{a(t)})_{a(t)\in\mathcal{M}_U} (e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}]$, and then it can be extended to a (∂_z, Δ_t)-differential field automorphism of $(\mathbf{K}_U)^{\wedge}$ keeping K_U invariant. Since $(K_U)^{\sim} \subset$

$(\mathbf{K}_U)^\wedge$, and since $(K_U)^\sim$ is stable under τ_α, the map τ_α induces an element of $\mathrm{Aut}^{\Delta_t}_{\partial_z}((K_U)^\sim|K_U)$.

The parametrized exponential torus (or simply, the exponential torus) is the subgroup of $\mathrm{Aut}^{\Delta_t}_{\partial_z}((K_U)^\sim|K_U)$ consisting of the τ_α, where α is a character of $\mathbf{E}_U$. Notice that the matrices of the exponential torus belongs to $\mathrm{GL}_m(\mathbb{C})$, while the coefficients of the matrix of $\hat{m}$ depend upon t.

Example 2.16. Let $t = (t_1, t_2)$ and let us consider

$$\partial_z \begin{pmatrix} Y_1(z,t) \\ Y_2(z,t) \end{pmatrix} = \begin{pmatrix} -t_1 z^{-2} & 0 \\ 0 & -t_2 z^{-2} \end{pmatrix} \begin{pmatrix} Y_1(z,t) \\ Y_2(z,t) \end{pmatrix},$$

which admits $\left(\begin{smallmatrix} e^{t_1/z} & 0 \\ 0 & e^{t_2/z} \end{smallmatrix}\right)$ as fundamental solution. The parametrized exponential torus and the parametrized differential Galois group are both equal to

$$\left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \;\middle|\; \alpha, \beta \in \mathbb{C}^* \right\}.$$

Remark that the unparametrized exponential torus (see p. 80 of [van der Put and Singer 2003]) and the unparametrized differential Galois group are isomorphic to $(\mathbb{C}^*)^2$ if and only if t_1 and t_2 are linearly independent over $\mathbb{Q}$. In particular, the matrices of the parametrized exponential torus evaluated at a specialized value (u, v) of the parameter are not always equal to the matrices of the unparametrized exponential torus of the system

$$\partial_z \begin{pmatrix} Y_1(z,u,v) \\ Y_2(z,u,v) \end{pmatrix} = \begin{pmatrix} -u z^{-2} & 0 \\ 0 & -v z^{-2} \end{pmatrix} \begin{pmatrix} Y_1(z,u,v) \\ Y_2(z,u,v) \end{pmatrix}.$$

This is a difference between the exponential torus and the two other generators of the parametrized differential Galois group: the monodromy and the Stokes operators (see Definition 2.18 below).

Lemma 2.17. *Let $d(t)$ be a singular direction of* $(*)$ (*see Section 1D*). *The Stokes matrix* $\mathrm{St}_{d(t)}$ *induces an element of* $\mathrm{Aut}^{\Delta_t}_{\partial_z}((K_U)^\sim|K_U)$.

Proof. Let us recall the construction of the Stokes matrices. Let $d(t)$ be a singular direction and let k_r be the biggest level of $(*)$. The assumption we have made on $\mathfrak{D}$ (see Section 1D) tells us that there exists $t \mapsto d^{\pm}(t)$, continuous in t, such that

$$d(t) - \frac{\pi}{2k_r} < d^-(t) < d(t) < d^+(t) < d(t) + \frac{\pi}{2k_r},$$

with no singular directions in $[d^-(t), d(t)[\,\cup\,]d(t), d^+(t)]$. From the construction of $\mathrm{St}_{d(t)}$, and Section 1C, we know that

$$H^{d^+(t)}(z,t) e^{L(t)\log(z)} e^{Q(z,t)} = H^{d^-(t)}(z) e^{L(t)\log(z)} e^{Q(z,t)}\, \mathrm{St}_{d(t)}\,.$$

By construction, the Stokes matrix induces the identity on K_U. To prove that the Stokes matrices are elements of $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$, we have to prove that the maps $i^{\pm}$ that send $\hat{H}(z,t)z^{L(t)}e(Q(z,t))$ to $H^{d^{\pm}(t)}(z,t)e^{L(t)\log(z)}e^{Q(z,t)}$ induce (∂_z, Δ_t)-field isomorphisms. From the unparametrized case (see Theorem 2, §6.4 of [Balser 1994]), and the relations satisfied by the symbols log, $(z^{a(t)})_{a(t)\in\mathcal{M}_U}$ and $(e(q(z,t)))_{q(z,t)\in\mathbf{E}_U}$ (see Section 1A), $i^{\pm}$ induce ∂_z-field isomorphisms.

We want now to prove that if $\hat{H}(z,t)$ admits $H^{d^{\pm}(t)}(z,t)$ as asymptotic sum in the direction $d^{\pm}(t)$, then $\partial_{t_i}\hat{H}(z,t)$ admits $\partial_{t_i}H^{d^{\pm}(t)}(z,t)$ as asymptotic sum in the direction $d^{\pm}(t)$ for all $i\leq n$. This is a consequence of Lemma 1.14 and the fact that we may assume that the $d^{\pm}(t)$ are locally constant. Hence $i^{\pm}$ commute with ∂_{t_i}, and $i^{\pm}$ induce (∂_z, Δ_t)-field isomorphisms. □

Definition 2.18. Let $d(t)$ be a singular direction of $(*)$. Then the element of $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$ induced by the Stokes matrix in the direction $d(t)$ is the Stokes operator in the direction $d(t)$. For simplicity of notation, we write $\mathrm{St}_{d(t)}$ for both the Stokes operator and the Stokes matrix in the direction $d(t)$.

Proposition 2.19. *If $g(z,t)\in(K_U)^{\sim}$ is fixed by all the Stokes operators $\mathrm{St}_{d(t)}$, the monodromy and the exponential torus, then $g(z,t)\in K_U$.*

Proof. Let $\overline{\mathcal{M}}_U$ be the algebraic closure of $\mathcal{M}_U$. Proposition 3.25 of [van der Put and Singer 2003] implies that if $g(z,t)\in(\mathbf{K}_U)^{\wedge}$ is fixed by the monodromy and the exponential torus, then $g(z,t)\in(\mathbf{K}_U)^{\wedge}\cap\overline{\mathcal{M}}_U[[z]][z^{-1}]=\hat{K}_U$. Since $(K_U)^{\sim}\subset(\mathbf{K}_U)^{\wedge}$, we have to prove that if $g(z,t)\in(K_U)^{\sim}\cap\hat{K}_U$ is fixed by all the Stokes operators, then $g(z,t)\in K_U$. Let $g(z,t)\in(K_U)^{\sim}\cap\hat{K}_U$ be fixed by all the Stokes operators. Let $F(z,t)=\hat{H}(z,t)z^{L(t)}e(Q(z,t))$ be the fundamental solution defined in Proposition 1.3, and let $(\hat{H}_{i,j})$ be the entries of the matrix $\hat{H}(z,t)$. There exists $P\in K_U\langle X_{i,j}\rangle_{\partial_z,\Delta_t}$ such that $P(\hat{H}_{i,j})=g(z,t)$. Let $d(t)$ satisfy the same properties as in Proposition 1.13. Because of Proposition 1.13, there exists a map $U\to\mathbb{R}^{>0}$, $t\mapsto\varepsilon(t)$ (which is not necessarily continuous) such that $P(H_{i,j}^{d(t)})$ is meromorphic in (z,t) for

$$z\in\Sigma(d_1(t)-\pi/2k_r,\, d_2(t)+\pi/2k_r)\text{ with }0<|z|<\varepsilon(t)\quad\text{and}\quad t\in U,$$

where $d_1(t), d_2(t)$ are two singular directions. Since $P(\hat{H}_{i,j})$ is fixed by all the Stokes operators, $P(H_{i,j}^{d(t)})$ is meromorphic in (z,t) for $0<|z|<\varepsilon(t)$ and $(z,t)\in\widetilde{\mathbb{C}}\times U$. Moreover, $P(H_{i,j}^{d(t)})(z,t)=P(H_{i,j}^{d(t)})(e^{2i\pi}z,t)$ on its domain of definition, which means that $P(H_{i,j}^{d(t)})$ is meromorphic in (z,t) for $0<|z|<\varepsilon(t)$ and $(z,t)\in\mathbb{C}\times U$. We recall that K_U consists of elements $f(z,t)\in\hat{K}_U$ such that for $0<|z|<\varepsilon(t)$, the function $t\mapsto f(z,t)$ lies in $\mathcal{M}_U$. We have $P(H_{i,j}^{d(t)})\in K_U$. We have seen in Lemma 2.17 that the map that sends $\hat{H}(z,t)z^{L(t)}e(Q(z,t))$ to $H^{d(t)}(z,t)e^{L(t)\log(z)}e^{Q(z,t)}$ induces a (∂_z,Δ_t)-field isomorphism. Since this map leaves K_U invariant, this implies that $P(\hat{H}_{i,j})=g(z,t)\in K_U$. □

We can now prove the main theorem of this paper. We recall some notation. Let Consider the equation $(*)$ with $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$ (see page 90), let K_U be the fraction field of $\mathbb{O}_U(\{z\})$, and let $(K_U)^{\sim}|K_U$ be the parametrized Picard–Vessiot extension defined in the beginning of Section 2D. Let $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$ be the field automorphisms of $(K_U)^{\sim}$ which commute with all the derivations and leave K_U invariant.

Theorem 2.20 (parametrized analogue of the density theorem of Ramis). *The group generated by the monodromy, the exponential torus and the Stokes operators is dense for the Kolchin topology in* $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$.

Proof. First of all, we have already pointed out that the monodromy, the exponential torus and the Stokes operators are elements of $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$. Using Proposition 2.9, we just have to prove that if $\alpha(z,t) \in (K_U)^{\sim}$ is fixed by the monodromy, the exponential torus and the Stokes operators, then it belongs to K_U. This is exactly Proposition 2.19. □

Remark 2.21. (1) Let $\mathbb{C}(t)\{z\}$ be the subset of $\mathbb{O}_U(\{z\})$ consisting of elements of the form $\sum_{i>N} a_i(t)z^i$, with $a_i(t) \in \mathbb{C}(t)$ and $N \in \mathbb{Z}$. Let us consider the equation $(*)$ with $A(z,t) \in \mathrm{M}_m(\mathbb{C}(t)\{z\})p$. Even if we were able to define a parametrized Picard–Vessiot extension over $\mathbb{C}(t)\{z\}$, we would not have a parametrized analogue of the density theorem of Ramis, because the monodromy is not defined in this case. In general, we have

$$\hat{m}(z^{\alpha(t)}) = e^{2i\pi\alpha(t)} z^{\alpha(t)} \notin \mathbb{C}(t)\{z\}(z^{\alpha(t)}).$$

This is why we take a larger field of constants with respect to ∂_z.

(2) Similarly, we can prove that the group generated by the monodromy and the exponential torus is dense for the Kolchin topology in $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|\hat{K}_U \cap (K_U)^{\sim})$.

Corollary 2.22. $\mathrm{Aut}_{\partial_z}^{\Delta_t}((K_U)^{\sim}|K_U)$ *contains a finitely generated Kolchin-dense subgroup.*

Proof. Let $q_1(z,t), \ldots, q_\beta(z,t) \in \mathbf{E}_U$ be $\mathbb{Q}$-linearly independent such that

$$(K_U)^{\sim} \subset \hat{K}_{F,U}(e(q_1(z,t)), \ldots, e(q_\beta(z,t))).$$

Let τ_i be an element of the exponential torus that fixes the $e(q_j(z,t))$ for $j \neq i$, and that sends $e(q_i(z,t))$ to $ae(q_i(z,t))$, with a not a root of unity.

By definition of the singular directions (see Section 1D), there exists $\nu \in \mathbb{N}^*$ such that there are finitely many singular directions modulo $2\nu\pi$. Let $d_1(t), \ldots, d_k(t)$ be continuous singular directions such that, if $d(t)$ is a singular direction, then $d(t)$ is equal to one of the $d_i(t)$ modulo $2\nu\pi$. Let $g(z,t) \in (K_U)^{\sim}$ be fixed by the monodromy, $\tau_1, \ldots, \tau_\beta$, and $\mathrm{St}_{d_1(t)}, \ldots, \mathrm{St}_{d_k(t)}$. Using Proposition 2.9(2), it is sufficient to prove that $g(z,t) \in K_U$.

We can write $g(z,t)$ as an element of

$$\hat{K}_{F,U}(e(q_1(z,t)),\dots,e(q_{\beta-1}(z,t)))(e(q_\beta(z,t))).$$

Since the $q_i(z,t)\in \mathbf{E}_U$ are $\mathbb{Q}$-linearly independent, we know by construction that the $e(Nq_\beta(z,t))$, with $N\in\mathbb{Z}$, are $\mathbb{C}$-linearly independent over

$$\hat{K}_{F,U}(e(q_1(z,t)),\dots,e(q_{\beta-1}(z,t))).$$

If we add the fact that $g(z,t)$ is fixed by τ_β, we obtain

$$g(z,t)\in\hat{K}_{F,U}(e(q_1(z,t)),\dots,e(q_{\beta-1}(z,t))).$$

We apply the same argument β times to conclude that $g(z,t)\in\hat{K}_{F,U}\cap(K_U)^{\sim}$. By the construction of the Stokes operators, we have that $\mathrm{St}_{d(t)}=\mathrm{Id}$ if and only if $\mathrm{St}_{2\nu\pi+d(t)}=\mathrm{Id}$, where $\nu\in\mathbb{N}^*$ has been defined in the proof. Proposition 2.19 allows us to conclude that $g(z,t)\in K_U$. □

2E. ***The density theorem for the global parametrized differential Galois group.*** In this subsection, we consider parametrized linear differential equations of the form $(*)$, with $A(z,t)\in\mathrm{M}_m(\mathcal{M}_U(z))$. We want to prove a density theorem for the global parametrized differential Galois group. The result in the unparametrized case is due to Ramis, and a proof can be found for instance in [Mitschi 1996, Proposition 1.3]. The parametrized singularities of $(*)$ (that is, the poles of $A(z,t)$ as a rational function in z, possibly including ∞) belong to the algebraic closure of $\mathcal{M}_U$. Because of Remark 1.1, after taking a smaller nonempty polydisc U, we may assume that the set of parametrized singularities belongs to $\mathcal{M}_U$. We will write "singularity" instead of "parametrized singularity" when no confusion is likely to arise. Let $S=\{\alpha_1(t),\dots,\alpha_k(t)\}\subset\mathbb{P}_1(\mathcal{M}_U)$ be the set of the singularities of $(*)$. For any singularity $\alpha(t)$ of this equation, we may define its levels and its set of singular directions by considering

$$\partial_z Y(z-\alpha(t),t)=A(z-\alpha(t),t)Y(z-\alpha(t),t)\quad\text{if }\infty\not\equiv\alpha(t)\in S$$

and

$$\partial_z Y(z^{-1},t)=A(z^{-1},t)Y(z^{-1},t)\quad\text{if }\infty\equiv\alpha(t)\in S.$$

Let $(d_{i,j}(t))$ be the singular directions of $\alpha_i(t)$. As in Section 1D, we define

$$\mathcal{D}_{\alpha_i(t)}=\{t\in U\mid \text{ there exist } j,j'\in\mathbb{N}\text{ such that } d_{i,j}\not\equiv d_{i,j'}\text{ and } d_{i,j}(t)=d_{i,j'}(t)\}.$$

From Lemma 1.12, all the $\mathcal{D}_{\alpha_i(t)}$ are closed sets with empty interior. After taking a smaller nonempty polydisc U, we may assume that:

- There exists $\varepsilon>0$ such that for all $t\in U$ and for all $i\neq j$,

$$|\alpha_i(t)-\alpha_j(t)|>\varepsilon.$$

- $\mathcal{D}_{\alpha_i(t)} = \varnothing$ for all $i \le k$.
- For all singularities of $(*)$, the levels are independent of t.
- For all $t_0 \in U$ and all singularities $\infty \not\equiv \alpha(t) \in S$, the singular directions of the equation $\partial_z Y(z-\alpha(t), t) = A(z-\alpha(t), t)Y(z-\alpha(t), t)$ evaluated at t_0 are equal to the singular directions of the specialized system $\partial_z Y(z-\alpha(t), t_0) = A(z-\alpha(t), t_0)Y(z-\alpha(t), t_0)$.
- Similarly, for all $t_0 \in U$, the singular directions of the equation $\partial_z Y(z^{-1}, t_0) = A(z^{-1}, t_0)Y(z^{-1}, t_0)$ evaluated at t_0 are equal to the singular directions of the specialized system $\partial_z Y(z^{-1}, t_0) = A(z^{-1}, t_0)Y(z^{-1}, t_0)$.
- Every entry of every z-coefficient of $A(z,t)$ is analytic on U.

Let $x_0(t) \in \mathcal{M}_U$ and let $\varepsilon > 0$ such that

$$|x_0(t) - \alpha_j(t)| > \varepsilon \quad \text{and} \quad |\alpha_i(t) - \alpha_j(t)| > \varepsilon \quad \text{for all } t \in U, i < j \le k.$$

For all $i \le k$ and all $t \in U$, we define $U_{\alpha_i(t)}$, the polydisc in the z-plane with center $\alpha_i(t)$ and with radius ε. Let $d_{\alpha_i}(t)$ be a continuous ray from $\alpha_i(t)$ in $U_{\alpha_i(t)}$, let $b_{\alpha_i}(t)$ be the continuous point of $d_{\alpha_i}(t)$ with $|b_{\alpha_i}(t) - \alpha_i(t)| = \varepsilon$, and let $\gamma_{\alpha_i}(t)$ be a continuous path in $\mathbb{P}_1(\mathcal{M}_U)$ from $x_0(t)$ to $b_{\alpha_i}(t)$ such that $|\gamma_{\alpha_i}(t) - \alpha_j(t)| > \varepsilon/2$ for all $t \in U$ and all $j \le k$. Analytic continuation of $F(z,t) = (F_{i,j})$, that is, a germ of solution at $x_0(t)$ with the path $\gamma_{\alpha_i}(t)$ and $d_{\alpha_i}(t)$, provides a fundamental solution $F^{d_{\alpha_i}(t)}(z,t)$ on a germ of open sector with vertex $\alpha_i(t)$ bisected by $d_{\alpha_i}(t)$.

Let $(\mathcal{M}_U(z))^{\sim} = \mathcal{M}_U(X)\langle F_{i,j}\rangle_{\partial_z,\Delta_t}$. From the assumptions we have made on $x_0(t)$, we deduce that this field has a field of constants with respect to ∂_z equal to $\mathcal{M}_U$. Therefore, we deduce that $(\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)$ is a parametrized Picard–Vessiot extension. The results of Section 2B may be applied here and we can define a parametrized differential Galois group $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$, which will be identified with a linear differential algebraic subgroup of $\mathrm{GL}_m(\mathcal{M}_U)$. We will make the same abuse of language as in the local case (see Remark 2.8) and call it the parametrized linear differential Galois group, or Galois group, if no confusion is likely to arise. As in Proposition 2.12, we want to prove now that it is the "same" as the one of Section 2A.

Let C be a (Δ_t)-differentially closed field that contains $\mathcal{M}_U$, and let $C(z)$ denote the (∂_z, Δ_t)-differential field of rational functions in the indeterminate z with coefficients in C, where z is a (Δ_t)-constant with $\partial_z z = 1$, C is the field of constants with respect to ∂_z, and ∂_z commutes with all the derivations. The next proposition is the analogue in the global case of Proposition 2.12.

Proposition 2.23. *Let us keep the same notation. Consider the equation* $\partial_z Y(z,t) = A(z,t)Y(z,t)$, *with* $A(z,t) \in \mathrm{M}_m(\mathcal{M}_U(z))$. *The extension field*

$$C(z)\langle F_{i,j}\rangle_{\partial_z,\Delta_t}|C(z) := (C(z))^{\sim}|C(z)$$

is a parametrized Picard–Vessiot extension for $\partial_z Y(z,t) = A(z,t)Y(z,t)$. *Moreover, there exist* $P_1, \dots, P_k \in \mathcal{M}_U\{X_{i,j}\}_{\Delta_t}$ *such that the image of the representation of* $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$ (*resp.* $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$) *associated to* $F(z,t)$ *is the set of* C-*rational points* (*resp.* $\mathcal{M}_U$-*rational points*) *of the linear differential algebraic subgroup of* $\mathrm{GL}_m(C)$ (*resp.* $\mathrm{GL}_m(\mathcal{M}_U)$) *defined by* $P_1, \dots, P_k$. *More explicitly*:

$$\begin{aligned}\{F^{-1}\varphi(F) \mid \varphi \in \mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))\}\\ = \{A = (a_{i,j}) \in \mathrm{GL}_m(C) \mid P_1(a_{i,j}) = \cdots = P_k(a_{i,j}) = 0\},\\ \{F^{-1}\varphi(F) \mid \varphi \in \mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))\}\\ = \{A = (a_{i,j}) \in \mathrm{GL}_m(\mathcal{M}_U) \mid P_1(a_{i,j}) = \cdots = P_k(a_{i,j}) = 0\}.\end{aligned}$$

Proof. This is exactly the same reasoning as in Proposition 2.12. □

We want to find topological generators for $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ for the Kolchin topology.

For $\alpha(t) \in \mathcal{M}_U$, let

$$K_{U,\alpha(t)} = \{f(z - \alpha(t), t) \mid f(z,t) \in K_U\}$$

and let

$$K_{U,\infty} = \{f(z^{-1}, t) \mid f(z,t) \in K_U\}.$$

Let $\alpha(t) \in S$ and let $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|K_{U,\alpha(t)} \cap (\mathcal{M}_U(z))^{\sim})$ be the local Galois group for the fundamental solution $F^{d_{\alpha}(t)}(z,t)$ described above. If we conjugate $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|K_{U,\alpha(t)} \cap (\mathcal{M}_U(z))^{\sim})$ by the differential isomorphism defined by analytic continuation of $F(z,t)$ described above, we get an injective morphism of linear differential algebraic groups

$$\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|K_{U,\alpha(t)} \cap (\mathcal{M}_U(z))^{\sim}) \hookrightarrow \mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)).$$

Using the maps $i^{\pm}$ defined in the proof of Lemma 2.17 and the injection above, we can define the monodromy, the exponential torus and the Stokes operators for any singularities in S as elements of

$$\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)).$$

Theorem 2.24 (global parametrized analogue of the density theorem of Ramis). *Consider the equation* $\partial_z Y(z,t) = A(z,t)Y(z,t)$, *where* $A(z,t) \in \mathrm{M}_m(\mathcal{M}_U(z))$. *For* $\alpha(t) \in S$, *let* $G_{\alpha(t)}$ *be the subgroup of*

$$\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|K_{U,\alpha(t)} \cap (\mathcal{M}_U(z))^{\sim})$$

generated by the monodromy, the exponential torus and the Stokes operators. Let G *be the subgroup of* $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ *generated by the* $G_{\alpha(t)}$ *with* $\alpha(t) \in S$.

Then G in dense for the Kolchin topology in

$$\mathrm{Aut}^{\Delta_t}_{\partial_z}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)).$$

Proof. We use (2) of Proposition 2.9. We have to prove that the subfield of $(\mathcal{M}_U(z))^{\sim}$ fixed by G is $\mathcal{M}_U(z)$. Let $f(z,t) \in (\mathcal{M}_U(z))^{\sim}$ be fixed by G. Then, by the same reasoning as in Proposition 2.19, it follows that $f(z,t)$ belongs to $K_{U,\alpha(t)}$ for $\alpha(t) \in S$. Therefore, we deduce that $f(z,t)$ is meromorphic in (z,t) on $\mathbb{P}_1(\mathbb{C}) \times U$ and has a finite number of poles in the z-plane for t fixed. Hence $f(z,t) \in \mathcal{M}_U(z)$. □

In particular, this generalizes Theorem 4.2 in [Mitschi and Singer 2012], which says that if the equation has only regular singular poles, then the group generated by the monodromy at each pole is dense for the Kolchin topology in $\mathrm{Aut}^{\Delta_t}_{\partial_z}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$.

Corollary 2.25. $\mathrm{Aut}^{\Delta_t}_{\partial_z}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ *contains a finitely generated Kolchin-dense subgroup.*

Proof. In the proof of Theorem 2.24, we see that the global parametrized differential Galois group is generated by all local parametrized differential Galois groups. Since there is a finite number of singularities, this is a consequence of Corollary 2.22. □

2F. *Examples.* In all the examples, we will compute the global parametrized differential Galois group. This means that the base field is $\mathcal{M}_U(z)$.

Example 2.26. Let us consider the equation $\partial_z Y(z,t) = (t/z)Y(z,t)$. This example was considered by direct computations in Example 2.5, but we will compute here $\mathrm{Aut}^{\Delta_t}_{\partial_z}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ using the parametrized density theorem. The fundamental solution is (z^t) and the parametrized Picard–Vessiot extension over $\mathcal{M}_U(z)$ is $\mathcal{M}_U(z, z^t, \log)$ (we want the extension to be closed under the derivations ∂_z and ∂_t). The exponential torus and the Stokes matrices are trivial. The monodromy sends z^t to $e^{2i\pi t}z^t$. The element $e^{2i\pi t}$ satisfies the differential equation

$$\partial_t\left(\frac{\partial_t e^{2i\pi t}}{e^{2i\pi t}}\right) = 0.$$

Therefore, the Kolchin closure of the monodromy is contained in

$$\left\{a \in \mathcal{M}_U \;\middle|\; \partial_t\left(\frac{\partial_t a}{a}\right)\right\} = \{ce^{bt} \mid b \in \mathbb{C}, c \in \mathbb{C}^*\}.$$

Conversely, the map $z^t \mapsto ce^{bt}z^t$ is an element of $\mathrm{Aut}^{\Delta_t}_{\partial_z}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$. Finally, viewed as a linear differential algebraic subgroup of $\mathrm{GL}_1(\mathcal{M}_U)$,

$$\begin{aligned}\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)) &\simeq \left\{a \in \mathcal{M}_U \,\middle|\, \partial_t\Big(\frac{\partial_t a}{a}\Big) = 0\right\} \\ &= \{a \in \mathcal{M}_U \mid a \neq 0 \text{ and } a\partial_t^2 a - (\partial_t a)^2 = 0\} \\ &\subseteq \mathrm{GL}_1(\mathcal{M}_U).\end{aligned}$$

Example 2.27 (parametrized Euler equation). Let $f(t)$ be an analytic function different from 0, and let us consider the equation

$$\partial_z^2 Y(z,t) + \Big(\frac{1}{z} - \frac{1}{f(t)z^2}\Big)\partial_z Y(z,t) + \frac{1}{f(t)z^3} Y(z,t) = 0,$$

which can be seen as a system:

$$\partial_z \begin{pmatrix} Y(z,t) \\ \partial_z Y(z,t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{-1}{f(t)z^3} & \frac{1}{f(t)z^2} - \frac{1}{z} \end{pmatrix} \begin{pmatrix} Y(z,t) \\ \partial_z Y(z,t) \end{pmatrix}.$$

If $f \equiv 1$, we recognize the Euler equation. A fundamental solution is

$$\begin{pmatrix} 1 & \hat{F}(z,t) \\ \frac{1}{f(t)z^2} & \partial_z \hat{F}(z,t) \end{pmatrix} \begin{pmatrix} e\big(\frac{-1}{f(t)z}\big) & 0 \\ 0 & 1 \end{pmatrix},$$

where $\hat{F}(z,t) = -\sum_{n\geq 0} n!(f(t)z)^{n+1}$. The only singularity is 0. The monodromy is trivial. Let τ be an element of the exponential torus. Then the image of the fundamental solution under τ is

$$\begin{pmatrix} 1 & \hat{F}(z,t) \\ \frac{1}{f(t)z^2} & \partial_z \hat{F}(z,t) \end{pmatrix} \begin{pmatrix} \alpha e\big(\frac{-1}{f(t)z}\big) & 0 \\ 0 & 1 \end{pmatrix},$$

with $\alpha \in \mathbb{C}^*$. Therefore, the matrices of the elements of the exponential torus are of the form $\mathrm{Diag}(\alpha, 1)$, with $\alpha \in \mathbb{C}^*$. The only level of the system is 1 and the singular directions are the $\arg(f(t)^{-1}) + 2k\pi$, with $k \in \mathbb{Z}$. As we have seen in Proposition 1.10, we can compute the Stokes matrix with the Laplace and the Borel transforms. It follows from the definition of the formal Borel transform that

$$\hat{\mathcal{B}}_1(\hat{F}(z,t)) \equiv \log(1 - f(t)z).$$

Let $0 < \varepsilon < \pi/2$ be such that there are no singular directions in

$$\big[\arg(f(t)^{-1}) - \varepsilon, \arg(f(t)^{-1})\big[\,\cup\, \big]\arg(f(t)^{-1}), \arg(f(t)^{-1}) + \varepsilon\big].$$

Then the following matrices are fundamental solutions:

$$\begin{pmatrix} 1 & \mathcal{L}_{1,\arg(f(t)^{-1})+\varepsilon}(\log(1 - f(t)z)) \\ \frac{1}{f(t)z^2} & \partial_z \mathcal{L}_{1,\arg(f(t)^{-1})+\varepsilon}(\log(1 - f(t)z)) \end{pmatrix} \begin{pmatrix} e^{\frac{-1}{f(t)z}} & 0 \\ 0 & 1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & \mathscr{L}_{1,\arg(f(t)^{-1})-\varepsilon}(\log(1-f(t)z)) \\ \frac{1}{f(t)z^2} & \partial_z \mathscr{L}_{1,\arg(f(t)^{-1})-\varepsilon}(\log(1-f(t)z)) \end{pmatrix} \begin{pmatrix} e^{\frac{-1}{f(t)z}} & 0 \\ 0 & 1 \end{pmatrix}.$$

To compute the Stokes matrix in the direction $\arg(f(t)^{-1})$, we have to compute

$$\mathscr{L}_{1,\arg(f(t)^{-1})+\varepsilon}(\log(1-f(t)z)) - \mathscr{L}_{1,\arg(f(t)^{-1})-\varepsilon}(\log(1-f(t)z)).$$

We have

$$\begin{aligned} &\mathscr{L}_{1,\arg(f(t)^{-1})+\varepsilon}(\log(1-f(t)z)) - \mathscr{L}_{1,\arg(f(t)^{-1})-\varepsilon}(\log(1-f(t)z)) \\ &\qquad = z^{-1}\int_0^{\infty i(\arg(f(t)^{-1})+\varepsilon)} \log(1-f(t)u)e^{-u/z}d(u) \\ &\qquad\qquad -z^{-1}\int_0^{\infty i(\arg(f(t)^{-1})-\varepsilon)} \log(1-f(t)u)e^{-u/z}d(u). \end{aligned}$$

Integration by parts and the residue theorem imply that

$$\mathscr{L}_{1,\arg(f(t)^{-1})+\varepsilon}(\log(1-f(t)z)) - \mathscr{L}_{1,\arg(f(t)^{-1})-\varepsilon}(\log(1-f(t)z)) = 2i\pi f(t)e^{-\frac{1}{f(t)z}}.$$

Therefore, the Stokes matrix in this direction is $\left(\begin{smallmatrix} 1 & 2i\pi f(t) \\ 0 & 1 \end{smallmatrix}\right)$. Finally, we obtain

$$\begin{aligned} \mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)) &\simeq \left\{ \begin{pmatrix} \alpha & bf \\ 0 & 1 \end{pmatrix} \,\middle|\, \alpha \in \mathbb{C}^* \text{ and } b \in \mathbb{C} \right\} \\ &\simeq \left\{ \begin{pmatrix} \alpha & \beta \\ 0 & 1 \end{pmatrix} \,\middle|\, \partial_t \alpha = 0, \alpha \neq 0 \text{ and } \partial_t\left(\frac{\beta}{f}\right) = 0 \right\}. \end{aligned}$$

Example 2.28 (Bessel equation). We are interested in the parametrized linear differential equation

$$\partial_z \begin{pmatrix} Y(z,t) \\ \partial_z Y(z,t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{t^2-z^2}{z^2} & \frac{-1}{z} \end{pmatrix} \begin{pmatrix} Y(z,t) \\ \partial_z Y(z,t) \end{pmatrix}.$$

This equation has two singularities: 0 and ∞. Let U be a nonempty disc such that $U \cap (1/2+\mathbb{Z}) = \varnothing$. First we will compute the local group at 0,

$$\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|K_{U,0} \cap (\mathcal{M}_U(z))^{\sim}).$$

If $t+1/2 \notin \mathbb{Z}$, the two solutions

$$J_t(z) = \left(\frac{z}{2}\right)^t \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{k!\Gamma(t+k+1)2^k},$$

$$J_{-t}(z) = \left(\frac{z}{2}\right)^{-t} \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{k!\Gamma(-t+k+1)2^k},$$

are linearly independent (see [Watson 1944, p. 43]) and we have a fundamental solution of the specialized system. The equation is regular singular at $z=0$, and therefore the group generated by the monodromy $\hat{m}$ is dense for the Kolchin topology in the parametrized differential Galois group $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|K_{U,0}\cap(\mathcal{M}_U(z))^{\sim})$. By the same reasoning as in Example 2.26,

$$\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|K_{U,0}\cap(\mathcal{M}_U(z))^{\sim}) \simeq \left\{\begin{pmatrix}\alpha & 0\\ 0 & \alpha^{-1}\end{pmatrix} \;\middle|\; \alpha\neq 0,\ \alpha\partial_t^2\alpha-(\partial_t\alpha)^2=0\right\}.$$

We now turn to the singularity at infinity. We have

$$\partial_z\begin{pmatrix}Y(z^{-1},t)\\ \partial_z Y(z^{-1},t)\end{pmatrix}=\begin{pmatrix}0 & 1\\ \frac{t^2}{z^2}-\frac{1}{z^4} & \frac{-1}{z}\end{pmatrix}\begin{pmatrix}Y(z^{-1},t)\\ \partial_z Y(z^{-1},t)\end{pmatrix}.$$

In order to compute the matrices of the monodromy, the elements of the exponential torus and the Stokes operators, we make use of another basis of solutions:

$$H_t^{(1)}(z^{-1})=\frac{J_{-t}(z^{-1})-e^{-it\pi}J_t(z^{-1})}{i\sin(t\pi)},$$
$$H_t^{(2)}(z^{-1})=\frac{J_{-t}(z^{-1})-e^{it\pi}J_t(z^{-1})}{-i\sin(t\pi)}.$$

In [Watson 1944, p. 198], we find that on the sector $]-\pi,2\pi[$, $H_t^{(1)}(z^{-1})$ is asymptotic to

$$\widetilde{H}_t^{(1)}(z^{-1})=\left(\frac{2z}{\pi}\right)^{1/2}e^{i(z^{-1}-t\pi/2-\pi/4)}\sum_{k=0}^{\infty}\frac{(-1)^k\Gamma(t+k+1/2)z^k}{(2i)^k k!\,\Gamma(t-k+1/2)}.$$

Similarly, on the sector $]-2\pi,\pi[$, $H_t^{(2)}(z^{-1})$ is asymptotic to

$$\widetilde{H}_t^{(2)}(z^{-1})=\left(\frac{2z}{\pi}\right)^{1/2}e^{-i(z^{-1}-t\pi/2-\pi/4)}\sum_{k=0}^{\infty}\frac{\Gamma(t+k+1/2)z^k}{(2i)^k k!\,\Gamma(t-k+1/2)}.$$

It follows that in the basis $(H_t^{(1)}(z^{-1}),H_t^{(2)}(z^{-1}))$, the matrix of the monodromy is

$$\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}$$

and the matrices of the elements of the exponential torus are of the form

$$\begin{pmatrix}\alpha & 0\\ 0 & \alpha^{-1}\end{pmatrix},\quad \text{where } \alpha\in\mathbb{C}^*.$$

The only level is 1, and due to the expression of $\widetilde{H}_t^1(z^{-1})$ and $\widetilde{H}_t^2(z^{-1})$, the singular directions are the directions $\pi/2+k\pi$, with $k\in\mathbb{Z}$. By definition, the Stokes matrix in the direction $\pi/2+k\pi$ is the matrix that sends the asymptotic representation defined on the sector $](k-1)\pi, (k+1)\pi[$ to the asymptotic representation defined on the sector $]k\pi, (k+2)\pi[$. In [Ramis and Martinet 1990, §3.4.12] (see also [Bertrand 1992]), we find that in the basis $(H_t^1(z^{-1}), H_t^2(z^{-1}))$ the Stokes matrix in the direction $\pi/2+2k\pi$ is

$$\begin{pmatrix} 1 & 0 \\ 2e^{2i\pi t}\cos(\pi t) & 1 \end{pmatrix},$$

and the Stokes matrix in the direction $-\pi/2+2k\pi$ is

$$\begin{pmatrix} 1 & -2e^{-2i\pi t}\cos(\pi t) \\ 0 & 1 \end{pmatrix}.$$

An application of the local and global density theorems (Theorems 2.24 and 2.20) gives that

$$\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim} | K_{U,\infty} \cap (\mathcal{M}_U(z))^{\sim}) \quad \text{and} \quad \mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim} | \mathcal{M}_U(z))$$

are linear differential algebraic subgroups of $\mathrm{SL}_2(\mathcal{M}_U)$, because all the matrices we have computed are in $\mathrm{SL}_2(\mathcal{M}_U)$, which is closed in the Kolchin topology.

Let C be a differentially closed field that contains $\mathcal{M}_U$, and consider the parametrized differential Galois group $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$ defined in Proposition 2.23. First, we are going to compute the Zariski closure G of $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$. Let $C^* = C \setminus \{0\}$. From the classification of linear algebraic subgroups of $\mathrm{SL}_2(C)$ (see [van der Put and Singer 2003, Theorem 4.29]), there are four possibilities:

(1) G is conjugate to a subgroup of $B = \left\{\left(\begin{smallmatrix} a & b \\ 0 & a^{-1}\end{smallmatrix}\right) \mid a \in C^*, b \in \mathbb{C}\right\}$.

(2) G is conjugate to a subgroup of $D_\infty = \left\{\left(\begin{smallmatrix} a & 0 \\ 0 & a^{-1}\end{smallmatrix}\right) \mid a \in C^*\right\} \cup \left\{\left(\begin{smallmatrix} 0 & b^{-1} \\ -b & 0\end{smallmatrix}\right) \mid b \in C^*\right\}$.

(3) G is finite.

(4) $G = \mathrm{SL}_2(C)$.

From Proposition 2.23, every matrix that belongs to $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ belongs also to $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$. Since G must contain

$$\begin{pmatrix} 1 & 0 \\ 2e^{2i\pi t}\cos(\pi t) & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & -2e^{-2i\pi t}\cos(\pi t) \\ 0 & 1 \end{pmatrix},$$

we find that the only possibility is that $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$ is Zariski-dense in $\mathrm{SL}_2(C)$. Cassidy [1972, Proposition 42] classified the Zariski-dense differential algebraic subgroups of $\mathrm{SL}_2(C)$. Finally, we have two possibilities:

- $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$ is conjugate to $\mathrm{SL}_2(C_0)$ over $\mathrm{SL}_2(C)$, where

$$C_0 = \{a \in C(z) \mid \partial_z a = \partial_t a = 0\}.$$

- $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z)) = \mathrm{SL}_2(C)$.

If $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$ is conjugate to $\mathrm{SL}_2(C_0)$ over $\mathrm{SL}_2(C)$, the matrix of the monodromy of the singularity 0 is conjugate to a matrix $M \in \mathrm{SL}_2(C_0)$ over $\mathrm{SL}_2(C)$. Similar matrices have the same eigenvalues, so the eigenvalues of M are $e^{2i\pi t}$ and $e^{-2i\pi t}$, which is not possible if M belongs to $\mathrm{SL}_2(C_0)$. Because of Proposition 2.23, we find that

$$\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)) = \mathrm{SL}_2(\mathcal{M}_U).$$

3. Applications

We now give three applications of parametrized differential Galois theory. In Section 3A, we deal with linear differential equations that are completely integrable (see Definition 3.1). It was proved in [CS] that an equation is completely integrable if and only if its parametrized differential Galois group is conjugate over a differentially closed field to a group of constant matrices. We use the global density theorem (Theorem 2.24) to prove that the equation is completely integrable if and only if there exists a fundamental solution such that the matrices of the topological generators for the Galois group appearing in the global density theorem are constant matrices. As a corollary, we deduce that the equation is completely integrable if and only if the matrices of the topological generators for the Galois group given in the parametrized density theorem are conjugate over $\mathrm{GL}_m(\mathcal{M}_U)$ to constant matrices. In Section 3B, we study an entry of a Stokes operator at the singularity at infinity of the equation

$$\partial_z^2 Y(z,t) = (z^3+t)Y(z,t).$$

In particular, we prove that it is not ∂_t-finite: it satisfies no parametrized linear differential equation. This partially answers a question by Sibuya. In Section 3C, we deal with the inverse problem in parametrized differential Galois theory. Let k be a so-called universal (Δ_t)-field (see Section 3B). We give a necessary condition for a linear differential algebraic subgroup of $\mathrm{GL}_m(k)$ to be the global parametrized differential Galois group for some equation having coefficients in $k(z)$. The corresponding sufficient condition was proved in [Mitschi and Singer 2012, Corollary 5.2].

3A. ***Completely integrable equations.*** In this subsection, we study completely integrable equations. See also [Gorchinskiy and Ovchinnikov 2013] for an approach from the point of view of differential Tannakian categories.

Definition 3.1. Let $A_0 \in \mathrm{M}_m(\mathcal{M}_U(z))$. We say that the linear differential equation $\partial_0 Y = A_0 Y$ is completely integrable if there exist $A_1, \dots, A_n \in \mathrm{M}_m(\mathcal{M}_U(z))$ such that, for all $0 \le i,\ j \le n$,

$$\partial_{t_i} A_j - \partial_{t_j} A_i = A_i A_j - A_j A_i,$$

with $\partial_{t_0} = \partial_z$.

Sibuya [1990, Theorem A.5.2.3] has shown that if the parametrized linear differential equation $(*)$ is regular singular, then it is isomonodromic (see page 89 for the definition) if and only if it is completely integrable. This result is not true in the irregular case. The main reason is the fact that there are more topological generators in the parametrized differential Galois group.

Proposition 3.2. *Let $A_0(z,t) \in \mathrm{M}_m(\mathcal{M}_U(z))$ and let $(\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)$ be the parametrized Picard–Vessiot extension for $\partial_z Y(z,t) = A_0(z,t)Y(z,t)$ defined in Section 2E. The linear differential equation $\partial_z Y(z,t) = A_0(z,t)Y(z,t)$ is completely integrable if and only if there is a fundamental solution $F(z,t)$ in $(\mathcal{M}_U(z))^{\sim}$ such that the images of the topological generators of $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ (see Theorem 2.24) with respect to the representation associated to $F(z,t)$ belong to $\mathrm{GL}_m(\mathbb{C})$.*

Proof. Let C be a differentially closed field that contains $\mathcal{M}_U$ and let us consider $C(z)$ as in Section 2E. Let $(C(z))^{\sim}|C(z)$ be the parametrized Picard–Vessiot extension for $\partial_z Y(z,t) = A_0(z,t)Y(z,t)$, and let $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$ be the parametrized differential Galois group defined in Section 2A. We recall that if we take a different fundamental solution in $(\mathcal{M}_U(z))^{\sim}$ to compute the Galois group, we obtain a conjugate linear differential algebraic subgroup of $\mathrm{GL}_m(\mathcal{M}_U)$.

Using the global density theorem (Theorem 2.24), we find that there exists a fundamental solution such that the matrices of the topological generators for the Galois group appearing in the global density theorem are constant if and only if $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ is conjugate over $\mathrm{GL}_m(\mathcal{M}_U)$ to a subgroup of $\mathrm{GL}_m(\mathbb{C})$. Using Proposition 2.23, we find that $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ is conjugate over $\mathrm{GL}_m(\mathcal{M}_U)$ to a subgroup of $\mathrm{GL}_m(\mathbb{C})$ if and only if $\mathrm{Gal}_{\partial_z}^{\Delta_t}((C(z))^{\sim}|C(z))$ is conjugate over $\mathrm{GL}_m(C)$ to a subgroup of $\mathrm{GL}_m(C_0)$, where

$$C_0 = \{a \in C(z) \mid \partial_z a = \partial_{t_1} a = \cdots = \partial_{t_n} a = 0\}.$$

Proposition 3.9 of [CS] says that this occurs if and only if there exist $A_1, \ldots, A_n \in \mathrm{M}_m(C(z))$ such that, for all $0 \leq i, j \leq n$,

$$\partial_{t_i} A_j - \partial_{t_j} A_i = A_i A_j - A_j A_i,$$

with $\partial_{t_0} = \partial_z$. To finish, we follow the proof of Proposition 1.24 in [Di Vizio and Hardouin 2012]. Let $0 < i \leq n$ and let us consider

$$\partial_z A_i - \partial_{t_i} A_0 = A_0 A_i - A_i A_0.$$

By clearing the denominators, we obtain that every entry of every z-coefficient of A_i satisfies a finite set of polynomial equations with coefficients in $\mathcal{M}_U$. Since the polynomial equations have a solution in C, they must have a solution in the algebraic closure of $\mathcal{M}_U$. Using Remark 1.1, we find a nonempty polydisc $U' \subset U$ such that all the A_i belong to $\mathrm{M}_m(\mathcal{M}_{U'}(z))$. This concludes the proof. □

In the proof of Proposition 3.2, we have proved:

Corollary 3.3. *Let $A(z,t) \in \mathrm{M}_m(\mathcal{M}_U(z))$. The equation $(*)$ is completely integrable if and only if the matrices of the topological generators for the Galois group appearing in Theorem 2.24 are conjugate over $\mathrm{GL}_m(\mathcal{M}_U)$ to constant matrices.*

Remark 3.4. This corollary improves Proposition 3.9 in [CS]. The conjugation occurs in a field that is not differentially closed. Furthermore, we do not need the entire parametrized differential Galois group to be conjugate to a group of constant matrices in order to deduce that the equation $(*)$ is completely integrable.

Gorchinskiy and Ovchinnikov [2013] studied completely integrable parametrized linear differential equations using differential Tannakian categories. In particular, they proved that the notion of integrability with respect to all the parameters is equivalent to the notion of integrability with respect to each parameter separately, which generalizes [Dreyfus 2013, Proposition 9]. Furthermore, they improve Proposition 3.9 in [CS] by avoiding the assumption that the field of constants is differentially closed.

3B. ***On the hypertranscendence of a Stokes matrix.*** In this subsection, we will study the parametrized linear differential equation

$$\partial_z^2 Y(z,t) = (z^3+t)Y(z,t). \tag{3-1}$$

Sibuya [1975, Chapter 2] showed that there exists a formal solution $y_0(z,t)$ which admits an asymptotic representation $\tilde{y}_0(z,t)$ on the sector

$$\Sigma(-3\pi/5, 3\pi/5)$$

(see [ibid., Theorem 6.1]). We easily check that, for $k \in \mathbb{Z}$,

$$y_k(z,t) = y_0(e^{-2ki\pi/5}z, e^{-6ki\pi/5}t)$$

is a solution of (3-1) which has the asymptotic representation

$$\tilde{y}_k(z,t) = \tilde{y}_0(e^{-2ki\pi/5}z, e^{-6ki\pi/5}t)$$

on the sector $S_{k-1} \cup \bar{S}_k \cup S_{k+1}$, where

$$S_k = \Sigma((2k-1)\pi/5, (2k+1)\pi/5)$$

and $\bar{S}_k$ is its closure.

The asymptotic representation $\tilde{y}_k(z,t)$ is bounded uniformly on each compact set in the t-plane as $|z|$ tends to infinity on the sector S_k, and tends to infinity uniformly on each compact set in the t-plane as $|z|$ tends to infinity on the sectors S_{k-1} and S_{k+1}. As we see in [ibid., p. 83], $y_{k+1}(z,t)$ and $y_{k+2}(z,t)$ are linearly

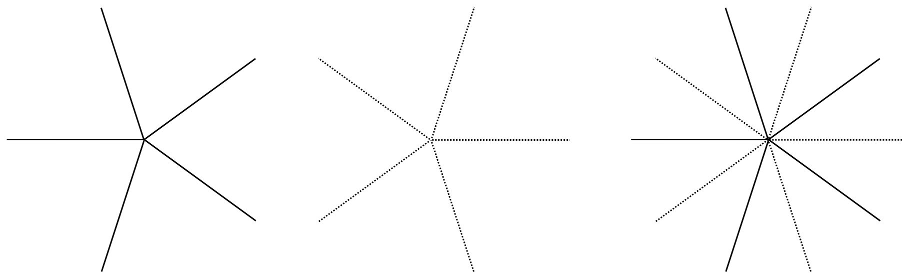

Figure 1. Left: the sectors S_k. Middle: the singular directions. Right: the sectors S_k and the singular directions.

independent, and we can write $y_k(z,t)$ as an $\mathcal{M}_{\mathbb{C}}$-linear combination of $y_{k+1}(z,t)$ and $y_{k+2}(z,t)$:

$$y_k(z,t) = C_k(t)y_{k+1}(z,t) + \widetilde{C}_k(t)y_{k+2}(z,t) \quad \text{for all } k \in \mathbb{N},\ z,t \in \mathbb{C}, \tag{3-2}$$

where $\widetilde{C}_k(t), C_k(t) \in \mathcal{M}_{\mathbb{C}}$. By Theorem 21.1 in [ibid.], we obtain that

$$\widetilde{C}_k(t) = -e^{2i\pi/5} \quad \text{and} \quad C_k(t) = C_0(e^{-6ki\pi/5}t).$$

Sibuya [1975] asked if $C_0(t)$ is differentially transcendental, i.e., satisfies no differential polynomial equations. We will use Galois theory to prove that for every nonempty polydisc U, $C_0(t)$ is not ∂_t-finite over $\mathcal{M}_U$, i.e., satisfies no linear differential equations in coefficients in $\mathcal{M}_U$.

The singularity of the system is at infinity. Let $W(z,t) = zY(z^{-1},t)$. We obtain the parametrized linear differential equation

$$z^7\partial_z^2 W(z,t) = (1+tz^3)W(z,t), \tag{3-3}$$

which can be written in the form

$$\partial_z \begin{pmatrix} W(z,t) \\ \partial_z W(z,t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{1+tz^3}{z^7} & 0 \end{pmatrix} \begin{pmatrix} W(z,t) \\ \partial_z W(z,t) \end{pmatrix}.$$

Let k be a so-called universal (Δ_t)-field of characteristic 0: for any (Δ_t)-field $k_0 \subset k$, (Δ_t)-finitely generated over $\mathbb{Q}$, and any (Δ_t)-finitely generated extension k_1 of k_0, there is a (Δ_t)-differential k_0-isomorphism of k_1 into k. See Chapter 3, §7 of [Kolchin 1973] for more details. In particular, k is (Δ_t)-differentially closed. Let $k(z)$ denote the (∂_z, Δ_t)-differential field of rational functions in the indeterminate z with coefficients in k, where z is a (Δ_t)-constant with $\partial_z z = 1$, k is the field of constants with respect to ∂_z, and ∂_z commutes with all the derivations.

Let

$$A(z,t) = \begin{pmatrix} 0 & 1 \\ \frac{1+tz^3}{z^7} & 0 \end{pmatrix}.$$

The two solutions $zy_1(z^{-1},t)$, $zy_2(z^{-1},t)$ admit asymptotic representations and the only singularity is 0. Therefore,

$$\mathcal{M}_U(z)\langle y_1(z^{-1},t), y_2(z^{-1},t)\rangle_{\partial_z,\partial_t}|\mathcal{M}_U(z) = (\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)$$

is a parametrized Picard–Vessiot extension for $\partial_z W(z,t) = A(z,t)W(z,t)$. By Proposition 2.23,

$$(k(z))^{\sim}|k(z) = k(z)\langle y_1(z^{-1},t), y_2(z^{-1},t)\rangle_{\partial_z,\partial_t}|k(z)$$

is a parametrized Picard–Vessiot extension.

Lemma 3.5. $$\mathrm{Gal}_{\partial_z}^{\Delta_t}((k(z))^{\sim}|k(z)) = \mathrm{SL}_2(k).$$

Notice that the differential equation is of the form $\partial_z^2 W(z,t) = r(z,t)W(z,t)$, where $r(z,t) \in k(z)$. In this case, we can compute the Galois group using a parametrized version of Kovacic's algorithm; see [Arreche 2012; Dreyfus 2013]. See also [Acosta-Humanez 2009; Acosta-Humánez et al. 2011]. In order to have a self contained proof, we will perform the calculations explicitly.

Proof. If we apply Kovacic's algorithm [1986], we find that the unparametrized differential Galois group $\mathrm{Gal}_{\partial_z}((k(z))^{\sim}|k(z))$ is equal to $\mathrm{SL}_2(k)$. We apply Proposition 6.26 in [Hardouin and Singer 2008], to deduce that $\mathrm{Gal}_{\partial_z}^{\Delta_t}((k(z))^{\sim}|k(z))$ is Zariski-dense in $\mathrm{SL}_2(k)$. By Proposition 42 in [Cassidy 1972], we deduce that there are two possibilities:

- $\mathrm{Gal}_{\partial_z}^{\Delta_t}((k(z))^{\sim}|k(z)) = \mathrm{SL}_2(k)$
- $\mathrm{Gal}_{\partial_z}^{\Delta_t}((k(z))^{\sim}|k(z))$ is conjugate to $\mathrm{SL}_2(k_0)$ over $\mathrm{SL}_2(k)$, where

$$k_0 = \{a \in k(z) \mid \partial_z a = \partial_t a = 0\}.$$

We see in [Dreyfus 2013, Remark 4.4] that the last case occurs if and only if the following parametrized differential equation has a solution in $\mathcal{M}_U(z)$, for some nonempty polydisc U in $\mathbb{C}^n$:

$$\partial_z^3 y(z,t) = \partial_z y(z,t)\frac{4+4tz^3}{z^7} + y(z,t)\partial_z\frac{4+4tz^3}{z^7} - \partial_t\frac{4+4tz^3}{z^7}.$$

With the algorithm presented in [van der Put and Singer 2003, p. 100], we find that this does not happen, so

$$\mathrm{Gal}_{\partial_z}^{\Delta_t}((k(z))^{\sim}|k(z)) = \mathrm{SL}_2(k). \qquad \square$$

Lemma 3.6. *The singular directions of* (3-3) *are*

$$\{2k\pi/5 \mid k \in \mathbb{Z}\}.$$

Proof. Let $k \in \mathbb{Z}$. The matrix

$$\begin{pmatrix} zy_k(z^{-1},t) & zy_{k+1}(z^{-1},t) \\ \partial_z zy_k(z^{-1},t) & \partial_z zy_{k+1}(z^{-1},t) \end{pmatrix}$$

is a fundamental solution for the equation

$$\partial_z \begin{pmatrix} W(z,t) \\ \partial_z W(z,t) \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{1+tz^3}{z^7} & 0 \end{pmatrix} \begin{pmatrix} W(z,t) \\ \partial_z W(z,t) \end{pmatrix}.$$

The fundamental solution admits an asymptotic representation on the sectors

$$\Sigma((2k-1)\pi/5, (2k+3)\pi/5).$$

The only level is $\frac{5}{2}$. From Proposition 1.13 and the construction of the singular directions, we find that the singular directions are $\{2k\pi/5 \mid k \in \mathbb{Z}\}$. □

Example 3.7. We want to compute the Stokes matrix in the direction $8\pi/5$ for the fundamental solution

$$\begin{pmatrix} zy_1(z^{-1},t) & zy_2(z^{-1},t) \\ \partial_z zy_1(z^{-1},t) & \partial_z zy_2(z^{-1},t) \end{pmatrix}.$$

We recall the construction of the Stokes matrices. See Section 1C for the notation. Let $\hat{H}(z,t)z^{L(t)}e(Q(z,t))$ be a fundamental solution in parametrized Hukuhara–Turrittin canonical form. Let $H^-(z,t)$ and $H^+(z,t))$ be the matrices such that

$$H^-(z,t)e^{L(t)\log(z)}e^{Q(z,t)} \quad \text{and} \quad H^+(z,t)e^{L(t)\log(z)}e^{Q(z,t)}$$

are the germs of asymptotic solutions on the sectors

$$\Sigma(\pi, 9\pi/5) \quad \text{and} \quad \Sigma(7\pi/5, 11\pi/5),$$

respectively. The Stokes matrix in the direction $8\pi/5$ is the matrix that sends

$$H^-(z,t)e^{L(t)\log(z)}e^{Q(z,t)} \quad \text{to} \quad H^+(z,t)e^{L(t)\log(z)}e^{Q(z,t)}.$$

With the domain of definition of the asymptotic representation of $z\tilde{y}_1(z^{-1},t)$, we deduce from the definition of the Stokes operators that

$$\mathrm{St}_{8\pi/5}(zy_1(z^{-1},t)) = zy_1(z^{-1},t). \tag{3-4}$$

We first write $\mathrm{St}_{8\pi/5}(zy_2(z^{-1},t))$ in the basis

$$(zy_0(z^{-1},t), zy_1(z^{-1},t)).$$

There exist $a(t)$ and $b(t) \in \mathcal{M}_U$ such that

$$\mathrm{St}_{8\pi/5}(zy_2(z^{-1},t)) = a(t)zy_0(z^{-1},t) + b(t)zy_1(z^{-1},t).$$

By the construction of the asymptotic solutions with Laplace and Borel transforms (see Proposition 1.10), the asymptotic representation of $\mathrm{St}_{8\pi/5}(zy_2(z^{-1},t))$ has to be bounded in some sector of $]7\pi/5, 11\pi/5[$, which means that there exist

$$\frac{7\pi}{5} < \alpha < \beta < \frac{11\pi}{5} \quad \text{and} \quad \varepsilon > 0$$

such that $\mathrm{St}_{8\pi/5}(zy_2(z^{-1},t))$ is uniformly bounded for $\arg(z) \in]\alpha, \beta[$ and $z < |\varepsilon|$. Therefore, $a(t) = 0$ or $b(t) = 0$. Since the Stokes operators are automorphisms, we get $b(t) = 0$. Lemma 3.5 says that the parametrized differential Galois group is $\mathrm{SL}_2(k)$. Therefore, because of Proposition 2.23 and Lemma 2.17, the determinant of the matrix has to be 1. Thus by (3-2), we get that the Stokes matrix in direction $8\pi/5$ is

$$\mathrm{St}_{8\pi/5} = \begin{pmatrix} 1 & -C_0(t)e^{3i\pi/5} \\ 0 & 1 \end{pmatrix}.$$

Lemma 3.8. *Let $C_0(t)$ be defined as above. Assume that $C_0(t)$ is ∂_t-finite over k. Then the ∂_t-differential transcendence degree* (*see Section 2A for definition*) *of $(k(z))^{\sim}$ over $k(z)$ is at most* 2.

Proof. The extension $(k(z))^{\sim}$ is generated over $k(z)$ by $y_1(z^{-1},t)$ and $y_2(z^{-1},t)$. By the parametrized differential Galois correspondence (see Theorem 9.5 in [CS]), the Kolchin closure of the group generated by $\mathrm{St}_{8\pi/5}$ is equal to

$$\mathrm{Gal}^{\Delta_t}_{\partial_z}((k(z))^{\sim}|F),$$

where F is the subfield of $(k(z))^{\sim}$ fixed by $\mathrm{St}_{8\pi/5}$. Using (3-4), we deduce that F contains

$$k(z)\langle y_1(z^{-1},t)\rangle_{\partial_z,\partial_t}.$$

Because $C_0(t)$ satisfies a linear differential equation with coefficients in k, there exists a linear differential polynomial P such that this group is of the form

$$\left\{\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \,\middle|\, P(\alpha) = 0 = P(C_0(t))\right\},$$

and has ∂_t-differential dimension over k equal to 0. Therefore by Proposition 2.7 the ∂_t-differential transcendence degree of $(k(z))^{\sim}$ over F is equal to 0. Since F contains $k(z)\langle y_1(z^{-1},t)\rangle_{\partial_z,\partial_t}$, there exists a differential polynomial Q with coefficients in $k(z)$ such that

$$Q(y_1(z^{-1},t), y_2(z^{-1},t)) = 0 = Q(\partial_z(y_1(z^{-1},t)), \partial_z(y_2(z^{-1},t))).$$

Therefore, the ∂_t-differential transcendence degree of $(k(z))^{\sim}$ over $k(z)$ is at most 2, because $(k(z))^{\sim}$ is generated as a ∂_t-differential field over $k(z)$ by

$$\{y_1(z^{-1},t), y_2(z^{-1},t), \partial_z(y_1(z^{-1},t)), \partial_z(y_2(z^{-1},t))\}. \qquad \square$$

Theorem 3.9. *The function $C_0(t)$ is not ∂_t-finite over k.*

Proof. As we see from Lemma 3.5,

$$\mathrm{Gal}_{\partial_z}^{\Delta_t}((k(z))^{\sim}|k(z)) = \mathrm{SL}_2(k).$$

Therefore, by Proposition 2.7, the ∂_t-differential transcendence degree of $(k(z))^{\sim}$ over $k(z)$ is 3. If $C_0(t)$ was ∂_t-finite over k, because of Lemma 3.8, the ∂_t-differential transcendence degree of $(k(z))^{\sim}$ over $k(z)$ would be smaller than 3. Therefore, $C_0(t)$ is not ∂_t-finite over k. □

3C. ***Which linear differential algebraic groups are parametrized differential Galois groups?*** As in Section 3B, let k be a universal (Δ_t)-field of characteristic 0. Let us consider the equation $(*)$ with $A(z,t) \in \mathrm{M}_m(k(z))$, let $(k(z))^{\sim}|k(z)$ be the parametrized Picard–Vessiot extension, and let

$$G = \mathrm{Gal}_{\partial_z}^{\Delta_t}((k(z))^{\sim}|k(z)) \subset \mathrm{GL}_m(k)$$

be the parametrized differential Galois group defined in Section 2A. The following theorem of Seidenberg, applied with $K_0 = \mathbb{Q}$ and K_1 the (Δ_t)-field generated by $\mathbb{Q}$ and the z-coefficients of $A(z,t)$, tells us that there exists a nonempty polydisc U such that $A(z,t)$ may be seen as an element of $\mathrm{M}_m(\mathcal{M}_U(z))$.

Theorem 3.10 [Seidenberg 1958; 1969]. *Let $\mathbb{Q} \subset K_0 \subset K_1$ be finitely generated (Δ_t)-differential extensions of $\mathbb{Q}$, and assume that K_0 consists of meromorphic functions on some domain U of $\mathbb{C}^n$. Then K_1 is isomorphic to the field K_1^* of meromorphic functions on a nonempty polydisc $U' \subset U$ such that $K_0|_{U'} \subset K_1^*$, and the derivations in Δ_t can be identified with the derivations with respect to the coordinates on U'.*

Let $(\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z)$ be the parametrized Picard–Vessiot extension defined in Section 2E and let $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ be the parametrized differential Galois group. Using Corollary 2.25, we find that $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$ contains a finitely generated subgroup that is Kolchin-dense in $\mathrm{Aut}_{\partial_z}^{\Delta_t}((\mathcal{M}_U(z))^{\sim}|\mathcal{M}_U(z))$. With Proposition 2.23, we find that G contains a finitely generated subgroup that is Kolchin-dense in G. Combined with Corollary 5.2 in [Mitschi and Singer 2012], which gives the sufficiency of the condition, this yields the following result:

Theorem 3.11 (inverse problem). *Let G be a linear differential algebraic subgroup of $\mathrm{GL}_m(k)$. Then G is the global parametrized differential Galois group of some equation having coefficients in $k(z)$ if and only if G contains a finitely generated subgroup that is Kolchin-dense in G.*

In the unparametrized case, any linear algebraic group defined over $\mathbb{C}$ is a Galois group of a Picard–Vessiot extension (see [Tretkoff and Tretkoff 1979]). In fact, every linear algebraic group defined over $\mathbb{C}$ contains a finitely generated subgroup that is Zariski-dense, which means that Theorem 3.11 is a generalization of the result in the previous reference.

The situation is more complicated in the parametrized case. For example, the additive group

$$\left\{ \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \,\middle|\, \alpha \in k \right\}$$

is not the global parametrized differential Galois group of any equation having coefficients in $k(z)$ (see Section 7 of [CS]). In the parametrized case with only regular singular poles, the problem was solved in [Mitschi and Singer 2012, Corollary 5.2]: they obtain the same necessary and sufficient condition on the group as in Theorem 3.11. Singer [2013] characterized the linear algebraic subgroups of $\mathrm{GL}_m(k)$ that appear as the global parametrized differential Galois groups of some equation having coefficients in $k(z)$: they are the groups such that the identity component has no quotient isomorphic to the additive group $(k, +)$ or multiplicative group $(k^*, \times)$ of k.

Appendix

Let us keep the same notation as in Section 1A and Section 1B. The goal of the appendix is to prove the following theorem. Notice that our proof closely follows the unparametrized case; see [Balser et al. 1980; Loday-Richaud 2001]. See Remark 1.6 for a discussion of another similar result.

Theorem A.1. *Consider the equation $\partial_z Y(z,t) = A(z,t)Y(z,t)$ with $A(z,t) \in \mathrm{M}_m(\hat{K}_U)$. There exists a nonempty polydisc $U' \subset U$ such that we have a fundamental solution of the form*

$$\hat{P}(z,t) z^{C(t)} e(Q(z,t)) \in \mathrm{GL}_m((\mathbf{K}_{U'})^\wedge),$$

with:

- $\hat{P}(z,t) \in \mathrm{GL}_m(\hat{K}_{U'})$,
- $C(t) \in \mathrm{M}_m(\mathcal{M}_{U'})$,
- $e(Q(z,t)) = \mathrm{Diag}(e(q_i(z,t)))$, *with* $q_i(z,t) \in \mathbf{E}_{U'}$.

Moreover, we may choose the same nonempty polydisc U' as in Proposition 1.3. Combined with Remark 1.6, if $A(z,t) \in \mathrm{M}_m(\mathbb{O}_U(\{z\}))$, this gives a sufficient condition on $t_0 \in U$ to have a fundamental solution $\hat{P}(z,t) z^{C(t)} e(Q(z,t)) \in \mathrm{GL}_m((\mathbf{K}_{U'})^\wedge)$ in the same form as above with $t_0 \in U'$.

Remark that, contrary to Proposition 1.3, $\hat{H}(z,t) \in \mathrm{GL}_m(\hat{K}_{U'})$. On the other hand, we lose the commutation between $z^{C(t)}$ and $e(Q(z,t))$. Before giving the proof of the theorem, we state and prove two lemmas.

Lemma A.2. *Let $U' \subset U$ be a nonempty polydisc. Let $a(t) \in \mathcal{M}_{U'}$ and $\alpha(z,t) \in \hat{K}_{F,U'}$ such that $\hat{m}(\alpha(z,t)) = a(t)\alpha(z,t)$. Then there exist $\hat{h}(z,t) \in \hat{K}_{U'}$ and $b(t) \in \mathcal{M}_{U'}$ such that $\alpha(z,t) = \hat{h}(z,t)z^{b(t)}$.*

Proof. Let $\alpha(z,t) \in \hat{K}_{F,U'}$ such that $\hat{m}(\alpha(z,t)) = a(t)\alpha(z,t)$. The element $\alpha(z,t)$ belongs to the fraction field of a free polynomial ring

$$P = \hat{K}_{U'}[\log, z^{b_1(t)}, \dots, z^{b_k(t)}].$$

Write $\alpha(z,t) = \alpha_1(z,t)/\alpha_2(z,t)$, where α_1 and α_2 have gcd 1 in P. Using the relations in $\hat{K}_{F,U'}$, and applying $\hat{m}$ to $\alpha_1(z,t)/\alpha_2(z,t)$, we find that $\alpha(z,t)$ contains no terms in log. One can normalize $\alpha_2(z,t)$ such that it contains a term of the form $z^{n_1 b_1(t)+\dots+n_k b_k(t)}$ with coefficient 1 and $n_i \in \mathbb{Z}$. Using

$$\hat{m}(\alpha_1(z,t)/\alpha_2(z,t)) = a(t)\alpha_1(z,t)/\alpha_2(z,t),$$

we find that

$$\hat{m}(\alpha_2(z,t)) = e^{2i\pi(n_1 b_1(t)+\dots+n_k b_k(t))}\alpha_2(z,t)$$

and

$$\hat{m}(\alpha_1(z,t)) = a(t)e^{2i\pi(n_1 b_1(t)+\dots+n_1 b_1(t))}\alpha_1(z,t),$$

which is impossible unless

$$e^{2i\pi(n_1 b_1(t)+\dots+n_k b_k(t))} = 1.$$

This means that $\alpha_2(z,t) \in \hat{K}_{U'}$ and we may assume $\alpha_2(z,t) = 1$. Applying $\hat{m}$ to $\alpha_1(z,t)$, one finds that $\alpha_1(z,t)$ contains at most one term, that is, $\alpha(z,t) = \hat{h}(z,t)z^{b(t)}$, with $\hat{h}(z,t) \in \hat{K}_{U'}$ and $b(t) \in \mathcal{M}_{U'}$ satisfying $e^{2i\pi b(t)} = a(t)$. $\square$

Lemma A.3. *Let $U' \subset U$ be a nonempty polydisc. Let $A(z,t) \in \mathrm{M}_m(\hat{K}_{U'})$. Let $F_1(z,t)e(Q_1(z,t))$ and $F_2(z,t)e(Q_2(z,t))$ be two fundamental solutions of the equation $(*)$ such that, for $i = 1, 2$, we have*

$$F_i(z,t) \in \mathrm{GL}_m(\hat{K}_{F,U'}) \quad \textit{and} \quad Q_i(z,t) = \mathrm{Diag}[q_{i,j}(z,t)],$$

where the $q_{i,j}(z,t)$ belong to $\mathbf{E}_{U'}$. Then $F_1(z,t)^{-1}F_2(z,t) \in \mathrm{GL}_m(\mathcal{M}_{U'})$.

Proof. A straightforward computation shows that

$$\partial_z g((F_1(z,t)e(Q_1(z,t)))^{-1}F_2(z,t)e(Q_2(z,t))g) = 0.$$

By Proposition 2.19,

$$(F_1(z,t)e(Q_1(z,t)))^{-1}F_2(z,t)e(Q_2(z,t)) = C(t) \in \mathrm{GL}_m(\mathcal{M}_{U'}).$$

Hence, we have the equality

$$e(Q_1(z,t))C(t)e(-Q_2(z,t)) = F_1(z,t)^{-1}F_2(z,t).$$

The entries of $e(Q_1(z,t))C(t)e(-Q_2(z,t))$ are of the form

$$C_{i,j}(t)e(q_{1,j}(z,t) - q_{2,j}(z,t)),$$

with $C_{i,j}(t) \in \mathcal{M}_{U'}$, and the matrix $F_1(z,t)^{-1}F_2(z,t)$ belongs to $\mathrm{GL}_m(\hat{K}_{F,U'})$. By construction, $\hat{K}_{F,U'} \cap \mathcal{M}_{U'}((e(q(z,t)))_{q(z,t)\in\mathbf{E}_{U'}}) = \mathcal{M}_{U'}$, and we obtain

$$F_1(z,t)^{-1}F_2(z,t) \in \mathrm{GL}_m(\mathcal{M}_{U'}). \qquad \square$$

Proof of Theorem A.1. By Proposition 1.3, we know that we have a fundamental solution of the parametrized linear differential equation $(*)$ of the form

$$\hat{H}(z,t)z^{L(t)}e(Q(z,t)),$$

with $\hat{H}(z,t) \in \mathrm{GL}_m(\hat{K}_{U'}[z^{1/\nu}])$ and $\nu \in \mathbb{N}^*$. From Definition 2.13, $\hat{m}$ commutes with the derivation ∂_z, and therefore $\hat{m}(\hat{H}(z,t)z^{L(t)}e(Q(z,t)))$ is another fundamental solution. From the construction of $\hat{m}$, we deduce that $\hat{m}(\hat{H}(z,t)z^{L(t)}) \in \mathrm{GL}_m(\hat{K}_{F,U'})$, and we can apply Lemma A.3 to deduce the existence of $\hat{M}(t) \in \mathrm{GL}_m(\mathcal{M}_{U'})$ such that

$$\hat{m}(\hat{H}(z,t)z^{L(t)}) = \hat{H}(z,t)z^{L(t)}\hat{M}(t). \tag{A-1}$$

Let us consider $\hat{M}(t) = D(t)U(t)$, with $D(t)$ diagonalizable and $U(t)$ unipotent such that $D(t)U(t) = U(t)D(t)$ is the multiplicative analogue of the Jordan decomposition of $\hat{M}(t)$. If $a(t)$ is an eigenvalue of $D(t)$ (and therefore an eigenvalue of $\hat{M}(t)$), then there exists $0 \neq \alpha(z,t) \in \hat{K}_{F,U'}$ such that $\hat{m}(\alpha(z,t)) = a(t)\alpha(z,t)$, because of the relation (A-1). By Lemma A.2, $\alpha(z,t)$ is equal to $\hat{h}(z,t)z^{b(t)}$, with $b(t) \in \mathcal{M}_{U'}$ satisfying $e^{2i\pi b(t)} = a(t)$ and $\hat{h}(z,t) \in \hat{K}_{U'}$. This implies that $a(t)$ and all the eigenvalues of $D(t)$ are of the form $e^{\beta(t)}$, with $\beta(t) \in \mathcal{M}_{U'}$. So we have proved the existence of $C(t) \in \mathrm{M}_m(\mathcal{M}_{U'})$ such that $e^{2i\pi C(t)} = \hat{M}(t)$. Let

$$\hat{P}(z,t) = \hat{H}(z,t)z^{L(t)}z^{-C(t)}.$$

A computation shows that the monodromy of $z^{C(t)}$ is

$$\hat{m}(z^{C(t)}) = e^{2i\pi C(t)}z^{C(t)} = z^{C(t)}e^{2i\pi C(t)}.$$

The matrix $\hat{P}(z,t)$ is fixed by the monodromy and therefore belongs to $\mathrm{GL}_m(\hat{K}_{U'})$, by Proposition 2.19. Finally,

$$\hat{P}(z,t)z^{C(t)}e(Q(z,t))$$

is a fundamental solution of the parametrized linear differential equation $(*)$ that has the required property. $\square$

Acknowledgements

This paper was prepared during my thesis, supported by the region Île de France. I want to thank my advisor Lucia Di Vizio for her helpful comments and the interesting discussions we had during the preparation of this paper. I also want to thank the organizers of the seminars that have made it possible for me to present the results contained in this paper. I want to thank Jean-Pierre Ramis, Guy Casale, Reinhard Schäfke, Daniel Bertrand and Michael F. Singer for pointing some mistakes and inaccuracies in this paper. Michael F. Singer in particular suggested the contribution of Theorem 3.11 to this paper. I certainly thank Carlos E. Arreche and Claude Mitschi for the read-through. Lastly, I heartily thank the anonymous referees of the two successive submissions and Jacques Sauloy, who spent a great lot of time and effort to help me make this paper readable.

References

[Acosta-Humanez 2009] P. B. Acosta-Humanez, *Galoisian Approach to Supersymmetric Quantum Mechanics, Phd Dissertation*, Ph.D. thesis, Universitat Politecnica de Catalunya, 2009.

[Acosta-Humánez et al. 2011] P. B. Acosta-Humánez, J. J. Morales-Ruiz, and J.-A. Weil, "Galoisian approach to integrability of Schrödinger equation", *Rep. Math. Phys.* **67**:3 (2011), 305–374. MR 2846216 Zbl 1238.81090

[Arreche 2012] C. E. Arreche, "Computing the differential galois group of a one-parameter family of second order linear differential equations", preprint, 2012. arXiv 1208.2226

[Babbitt and Varadarajan 1985] D. G. Babbitt and V. S. Varadarajan, "Deformations of nilpotent matrices over rings and reduction of analytic families of meromorphic differential equations", 325 (1985), iv+147. MR 87i:12014 Zbl 0583.34007

[Balser 1994] W. Balser, *From divergent power series to analytic functions*, Lecture Notes in Mathematics **1582**, Springer, Berlin, 1994. MR 96d:34071 Zbl 0810.34046

[Balser et al. 1980] W. Balser, W. B. Jurkat, and D. A. Lutz, "A general theory of invariants for meromorphic differential equations, III: Applications", *Houston J. Math.* **6**:2 (1980), 149–189. MR 83m:34003c Zbl 0506.34006

[Bertrand 1992] D. Bertrand, "Groupes algébriques et équations différentielles linéaires", 206 (1992), 183–204. MR 94b:34006 Zbl 0813.12004

[Bolibruch 1997] A. A. Bolibruch, "On isomonodromic deformations of Fuchsian systems", *J. Dynam. Control Systems* **3**:4 (1997), 589–604. MR 99c:34003 Zbl 0943.34083

[Cano and Ramis 1995] J. Cano and J.-P. Ramis, "Théorie de Galois différentielle, multisommabilité et phénomène de Stokes", notes from the course *Journées Galois Différentielles* conducted in May 1993 at Toulouse by J.-P. Ramis and M. Loday-Richaud, 1995, http://www.math.univ-toulouse.fr/~ramis/Cano-Ramis-Galois.pdf.

[Cassidy 1972] P. J. Cassidy, "Differential algebraic groups", *Amer. J. Math.* **94** (1972), 891–954. MR 50 #13058 Zbl 0258.14013

[Cassidy 1989] P. J. Cassidy, "The classification of the semisimple differential algebraic groups and the linear semisimple differential algebraic Lie algebras", *J. Algebra* **121**:1 (1989), 169–238. MR 90g:12007 Zbl 0678.14011

[Cassidy and Singer 2007] P. J. Cassidy and M. F. Singer, "Galois theory of parameterized differential equations and linear differential algebraic groups", pp. 113–155 in *Differential equations and quantum groups*, IRMA Lect. Math. Theor. Phys. **9**, Eur. Math. Soc., Zürich, 2007. MR 2008f:12010 Zbl 1230.12003

[Chatzidakis et al. 2008] 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, I*, edited by Z. Chatzidakis et al., London Math. Soc. Lecture Note Ser. **349**, Cambridge Univ. Press, 2008. MR 2009j:12014 Zbl 1234.12005

[Di Vizio and Hardouin 2012] L. Di Vizio and C. Hardouin, "Descent for differential Galois theory of difference equations: confluence and q-dependence", *Pacific J. Math.* **256**:1 (2012), 79–104. MR 2928542 Zbl 1258.12004

[Dreyfus 2013] T. Dreyfus, "Computing the Galois group of some parameterized linear differential equation of order two", preprint, 2013. To appear in *Proc. Amer. Math. Soc.* arXiv 1110.1053

[Écalle 1981] J. Écalle, *Les fonctions résurgentes, I*, Publications Mathématiques d'Orsay **81-05**, Université de Paris-Sud Département de Mathématique, Orsay, 1981. MR 84h:30077a Zbl 0499.30034

[Gillet et al. 2013] H. Gillet, S. Gorchinskiy, and A. Ovchinnikov, "Parameterized Picard–Vessiot extensions and Atiyah extensions", *Adv. Math.* **238** (2013), 322–411. MR 3033637

[Gorchinskiy and Ovchinnikov 2013] S. Gorchinskiy and A. Ovchinnikov, "Isomonodromic differential equations and differential tannakian categories", preprint, 2013.

[Hardouin and Singer 2008] C. Hardouin and M. F. Singer, "Differential Galois theory of linear difference equations", *Math. Ann.* **342**:2 (2008), 333–377. MR 2009j:39001 Zbl 1163.12002

[Kolchin 1973] E. R. Kolchin, *Differential algebra and algebraic groups*, Pure and Applied Mathematics **54**, Academic Press, New York, 1973. MR 58 #27929 Zbl 0264.12102

[Kolchin 1985] E. R. Kolchin, *Differential algebraic groups*, Pure and Applied Mathematics **114**, Academic Press, Orlando, FL, 1985. MR 87i:12016 Zbl 0556.12006

[Kovacic 1986] J. J. Kovacic, "An algorithm for solving second order linear homogeneous differential equations", *J. Symbolic Comput.* **2**:1 (1986), 3–43. MR 88c:12011 Zbl 0603.68035

[Landesman 2008] P. Landesman, "Generalized differential Galois theory", *Trans. Amer. Math. Soc.* **360**:8 (2008), 4441–4495. MR 2009i:12005 Zbl 1151.12004

[Loday-Richaud 1990] M. Loday-Richaud, "Introduction à la multisommabilité", *Gaz. Math.* 44 (1990), 41–63. MR 91h:40007 Zbl 0722.34005

[Loday-Richaud 1994] M. Loday-Richaud, "Stokes phenomenon, multisummability and differential Galois groups", *Ann. Inst. Fourier* (*Grenoble*) **44**:3 (1994), 849–906. MR 95g:34010 Zbl 0812.34004

[Loday-Richaud 1995] M. Loday-Richaud, "Solutions formelles des systèmes différentiels linéaires méromorphes et sommation", *Exposition. Math.* **13**:2-3 (1995), 116–162. MR 96i:34124 Zbl 0831.34002

[Loday-Richaud 2001] M. Loday-Richaud, "Rank reduction, normal forms and Stokes matrices", *Expo. Math.* **19**:3 (2001), 229–250. MR 2002k:34173 Zbl 0990.34076

[Loday-Richaud and Remy 2011] M. Loday-Richaud and P. Remy, "Resurgence, Stokes phenomenon and alien derivatives for level-one linear differential systems", *J. Differential Equations* **250**:3 (2011), 1591–1630. MR 2011m:34264 Zbl 1214.34087

[Magid 1994] A. R. Magid, *Lectures on differential Galois theory*, University Lecture Series **7**, Amer. Math. Soc., Providence, RI, 1994. MR 95j:12008 Zbl 0855.12001

[Malgrange 1983] B. Malgrange, "Sur les déformations isomonodromiques, II: Singularités irrégulières", pp. 427–438 in *Mathematics and physics* (Paris, 1979/1982), edited by L. Boutet de Monvel et al., Progr. Math. **37**, Birkhäuser, Boston, MA, 1983. MR 85m:58094b Zbl 0528.32018

[Malgrange 1991] B. Malgrange, *Équations différentielles à coefficients polynomiaux*, Progress in Mathematics **96**, Birkhäuser, Boston, MA, 1991. MR 92k:32020 Zbl 0764.32001

[Malgrange 1995] B. Malgrange, "Sommation des séries divergentes", *Exposition. Math.* **13**:2-3 (1995), 163–222. MR 96i:34125 Zbl 0836.40004

[Malgrange and Ramis 1992] B. Malgrange and J.-P. Ramis, "Fonctions multisommables", *Ann. Inst. Fourier* (*Grenoble*) **42**:1-2 (1992), 353–368. MR 93e:40007 Zbl 0759.34007

[Minchenko and Ovchinnikov 2011] A. Minchenko and A. Ovchinnikov, "Zariski closures of reductive linear differential algebraic groups", *Adv. Math.* **227**:3 (2011), 1195–1224. MR 2012k:12010 Zbl 1215.12009

[Mitschi 1996] C. Mitschi, "Differential Galois groups of confluent generalized hypergeometric equations: An approach using Stokes multipliers", *Pacific J. Math.* **176**:2 (1996), 365–405. MR 98f:12005 Zbl 0883.12004

[Mitschi and Singer 2012] C. Mitschi and M. F. Singer, "Monodromy groups of parameterized linear differential equations with regular singularities", *Bull. Lond. Math. Soc.* **44**:5 (2012), 913–930. MR 2975151 Zbl 1254.34124

[Mitschi and Singer 2013] C. Mitschi and M. F. Singer, "Projective isomonodromy and Galois groups", *Proc. Amer. Math. Soc.* **141**:2 (2013), 605–617. MR 2996965 Zbl 1268.34187

[Peón Nieto 2011] A. Peón Nieto, "On $\sigma\delta$-Picard–Vessiot extensions", *Comm. Algebra* **39**:4 (2011), 1242–1249. MR 2012c:12011 Zbl 1272.12018

[van der Put and Singer 2003] M. van der Put and M. F. Singer, *Galois theory of linear differential equations*, Grundlehren der Mathematischen Wissenschaften **328**, Springer, Berlin, 2003. MR 2004c:12010 Zbl 1036.12008

[Ramis 1980] J.-P. Ramis, "Les séries k-sommables et leurs applications", pp. 178–199 in *Complex analysis, microlocal calculus and relativistic quantum theory* (Les Houches, 1979), edited by D. Iagolnitzer, Lecture Notes in Phys. **126**, Springer, Berlin, 1980. MR 82k:32033 Zbl 1251.32008

[Ramis 1985] J.-P. Ramis, "Phénomène de Stokes et filtration Gevrey sur le groupe de Picard–Vessiot", *C. R. Acad. Sci. Paris Sér. I Math.* **301**:5 (1985), 165–167. MR 86k:12012 Zbl 0593.12015

[Ramis and Martinet 1990] J.-P. Ramis and J. Martinet, "Théorie de Galois différentielle et resommation", pp. 117–214 in *Computer algebra and differential equations*, edited by E. Tournier, Academic Press, London, 1990. MR 91d:12014 Zbl 0722.12007

[Ramis and Sibuya 1989] J.-P. Ramis and Y. Sibuya, "Hukuhara domains and fundamental existence and uniqueness theorems for asymptotic solutions of Gevrey type", *Asymptotic Anal.* **2**:1 (1989), 39–94. MR 90k:58209 Zbl 0699.34058

[Rasoamanana 2010] J.-M. Rasoamanana, "Résurgence-sommabilité de séries formelles ramifiées dépendant d'un paramètre et solutions d'équations différentielles linéaires", *Ann. Fac. Sci. Toulouse Math.* (6) **19**:2 (2010), 303–343. MR 2011g:34207 Zbl 1205.34122

[Remy 2012] P. Remy, "Matrices de Stokes–Ramis et constantes de connexion pour les systèmes différentiels linéaires de niveau unique", *Ann. Fac. Sci. Toulouse Math.* (6) **21**:1 (2012), 93–150. MR 2954106 Zbl 1244.34109

[Robinson 1959] A. Robinson, "On the concept of a differentially closed field", *Bull. Res. Council Israel Sect. F* **8F** (1959), 113–128. MR 23 #A2323 Zbl 0221.12054

[Schäfke 2001] R. Schäfke, "Formal fundamental solutions of irregular singular differential equations depending upon parameters", *J. Dynam. Control Systems* **7**:4 (2001), 501–533. MR 2002g:34199 Zbl 1029.34077

[Seidenberg 1958] A. Seidenberg, "Abstract differential algebra and the analytic case", *Proc. Amer. Math. Soc.* **9** (1958), 159–164. MR 20 #178 Zbl 0186.07502

[Seidenberg 1969] A. Seidenberg, "Abstract differential algebra and the analytic case, II", *Proc. Amer. Math. Soc.* **23** (1969), 689–691. MR 40 #1376 Zbl 0186.07503

[Sibuya 1975] Y. Sibuya, *Global theory of a second order linear ordinary differential equation with a polynomial coefficient*, North-Holland Mathematics Studies **18**, North-Holland, Amsterdam, 1975. MR 58 #6561 Zbl 0322.34006

[Sibuya 1990] Y. Sibuya, *Linear differential equations in the complex domain: Problems of analytic continuation*, Translations of Mathematical Monographs **82**, Amer. Math. Soc., Providence, RI, 1990. MR 92a:34010 Zbl 1145.34378

[Singer 2009] M. F. Singer, "Introduction to the Galois theory of linear differential equations", pp. 1–82 in *Algebraic theory of differential equations* (Edinburgh, 2006), edited by M. A. H. MacCallum and A. V. Mikhailov, London Math. Soc. Lecture Note Ser. **357**, Cambridge Univ. Press, 2009. MR 2010d:12006 Zbl 1176.12005

[Singer 2013] M. F. Singer, "Linear algebraic groups as parameterized Picard–Vessiot Galois groups", *J. Algebra* **373** (2013), 153–161. MR 2995020 Zbl 06182880

[Tretkoff and Tretkoff 1979] C. Tretkoff and M. Tretkoff, "Solution of the inverse problem of differential Galois theory in the classical case", *Amer. J. Math.* **101**:6 (1979), 1327–1332. MR 80k:12033 Zbl 0423.12021

[Umemura 1996] H. Umemura, "Galois theory of algebraic and differential equations", *Nagoya Math. J.* **144** (1996), 1–58. MR 98c:12009 Zbl 0885.12004

[Wasow 1965] W. Wasow, *Asymptotic expansions for ordinary differential equations*, Pure and Applied Mathematics **14**, Interscience, New York, 1965. Reprinted Dover, New York, 1987. MR 34 #3041 Zbl 0133.35301

[Watson 1944] G. N. Watson, *A treatise on the theory of Bessel functions*, 2nd ed., Cambridge University Press, Cambridge, England, 1944. Reprinted Cambridge, 1995. MR 6,64a Zbl 0174.36202

[Wibmer 2012] M. Wibmer, "Existence of ∂-parameterized Picard–Vessiot extensions over fields with algebraically closed constants", *J. Algebra* **361** (2012), 163–171. MR 2921616 Zbl 06125724

Received May 21, 2013.

THOMAS DREYFUS
INSTITUT DE MATHÉMATIQUES DE JUSSIEU
UNIVERSITÉ PARIS DIDEROT
4, PLACE JUSSIEU
75005 PARIS
FRANCE
thomas.dreyfus@imj-prg.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.143

ON THE CLASSIFICATION OF COMPLETE AREA-STATIONARY AND STABLE SURFACES IN THE SUBRIEMANNIAN SOL MANIFOLD

MATTEO GALLI

We study the classification of area-stationary and stable C^2 regular surfaces in the space of the rigid motions of the Minkowski plane $E(1,1)$, equipped with its subriemannian structure. We construct examples of area-stationary surfaces that are not foliated by subriemannian geodesics. We also prove that there exist an infinite number of C^2 area-stationary surfaces with a singular curve. Finally we show the stability of C^2 area-stationary surfaces foliated by subriemannian geodesics.

1. Introduction

The study of the subriemannian area functional in three-dimensional pseudohermitian manifolds and in other subriemannian spaces has been largely investigated in the last years, see [Ambrosio et al. 2006; Barbieri and Citti 2011; Barone Adesi et al. 2007; Bigolin and Cassano 2010; Capogna et al. 2009; Cheng and Hwang 2010; Cheng et al. 2012; 2005; 2007; Danielli et al. 2007; 2008; 2009; Galli 2013; Galli and Ritoré 2013; Garofalo and Nhieu 1996; Hladky and Pauls 2008; Hurtado et al. 2010; Hurtado and Rosales 2008; $\geq$ 2014; Ritoré 2009; Ritoré and Rosales 2008; Rosales 2012; Shcherbakova 2009], among others.

One of the more interesting questions concerning the subriemannian area functional is this:

Problem 1. Which are the area-minimizing surfaces in a given three-dimensional contact subriemannian manifold?

A surface Σ is *area-minimizing* if $A(\Sigma) \leqslant A(\tilde{\Sigma})$, for any compact deformation $\tilde{\Sigma}$ of Σ. To answer the previous question, a natural preliminary step is to study the *area-stationary* surfaces, the critical points of the area functional.

Research supported by MCyT-Feder grant MTM2010-21206-C02-01 and J. A. grant P09-FMQ-5088.
MSC2010: 49Q05, 49Q20, 53C17.
Keywords: subriemannian geometry, area-stationary surfaces, stable surfaces, pseudohermitian manifolds, Sol geometry.

Problem 2. Which are the area-stationary surfaces in a given three-dimensional contact subriemannian manifold?

We will consider these questions in the class of C^2 regular surfaces. For a general introduction about the study of the area functional in subriemannian spaces, we refer the interested reader to [Capogna et al. 2007] and [Galli 2012], which treat the case of $\mathbb{H}^n$ and the contact subriemannian manifolds respectively.

In Sasakian space forms, the classification of C^2 area stationary surfaces was given in [Hurtado et al. 2010] in the case of the Heisenberg group $\mathbb{H}^1$ and in [Rosales 2012] for the Sasakian structures of S^3 and $\widetilde{\mathrm{SL}}_2(\mathbb{R})$. In the case of pseudohermitian three-manifolds that are not Sasakian, the only known results concerning Problem 1 and Problem 2 are given in [Galli 2013], where the group of the rigid motions of the Euclidean plane $E(2)$ is studied.

Concerning the three-dimensional pseudohermitian manifolds, we have the following classification result, [Perrone 1998, Theorem 3.1], in terms of the Webster scalar curvature W and of the pseudohermitian torsion τ:

Proposition 1.1. *Let M be a simply connected contact 3-manifold that is homogeneous* (in the sense of [Boothby and Wang 1958]). *The following possibilities arise*:

1. *If M is unimodular, it is*
   - (i) *the first Heisenberg group $\mathbb{H}^1$ when $W = |\tau| = 0$;*
   - (ii) *the three-sphere group* SU(2) *when $W > 2|\tau|$;*
   - (iii) *the group $\widetilde{\mathrm{SL}}_2(\mathbb{R})$ when $-2|\tau| \neq W < 2|\tau|$;*
   - (iv) *the group $\widetilde{E(2)}$, universal cover of the group of rigid motions of the Euclidean plane, when $W = 2|\tau| > 0$;*
   - (v) *the group $E(1, 1)$ of rigid motions of Minkowski 2-space, when $W = -2|\tau| < 0$;*
2. *If M is not unimodular, the Lie algebra is given by*

$$[X, Y] = \alpha Y + 2T, \quad [X, T] = \gamma Y, \quad [Y, T] = 0, \quad \alpha \neq 0,$$

   where $\{X, Y\}$ is an orthonormal basis of $\mathcal{H}$, $J(X) = Y$ and T is the Reeb vector field. In this case $W < 2|\tau|$ and when $\gamma = 0$ the structure is Sasakian and $W = -\alpha^2$.

The only case for which Problems 1 and 2 have not been investigated is that of Sol geometry, modeled by the space $E(1, 1)$. Its study is the aim of the present work.

After some preliminaries, the paper is organized as follows.

In Section 3, we compute explicitly the coordinates of the characteristic curves with given initial conditions. These curves play an important role in the study of area-stationary surfaces, since the regular part $\Sigma - \Sigma_0$ of a surface Σ is foliated by characteristic curves, that are not in general subriemannian geodesics, since $E(1, 1)$ is characterized by a nonvanishing pseudohermitian torsion.

Section 4 is the core of the paper. We first characterize the C^2 complete, area-stationary surfaces immersed in $E(1, 1)$ with singular points or singular curves that are subriemannian geodesics. On the other hand, for the first time in the three-dimensional pseudohermitian setting, we also find examples of area-stationary surfaces that are not foliated by subriemannian geodesics. We stress that these examples form an infinite family; that is, given an horizontal curve Γ, we can construct an area-stationary surface having Γ as singular set Σ_0.

Finally in Section 5 we prove that complete area-stationary surfaces with nonempty singular set, whose characteristic curves are subriemannian geodesics, are stable. We also find three families of nonsingular planes that are area-minimizing, using a calibration argument.

We remark that Section 5 opens two interesting questions. Is a stable complete area-stationary surface in $E(1, 1)$ with a singular curve always foliated by subriemannian geodesics in $\Sigma - \Sigma_0$? Do some other complete stable area-stationary surfaces in $E(1, 1)$ with empty singular set exist?

2. Preliminaries

The group $E(1, 1)$ of rigid motions of the Minkowski plane. We consider the group of rigid motions of the Minkowski plane $E(1, 1)$, a unimodular Lie group with a natural subriemannian structure. As a model of $E(1, 1)$ we choose the underlying manifold $\mathbb{R}^3$ with the following orthonormal basis of left-invariant vector fields:

$$X = \frac{\partial}{\partial z}, \quad Y = \frac{1}{\sqrt{2}}\left(-e^z \frac{\partial}{\partial x} + e^{-z}\frac{\partial}{\partial y}\right), \quad T = \frac{1}{\sqrt{2}}\left(e^z \frac{\partial}{\partial x} + e^{-z}\frac{\partial}{\partial y}\right). \tag{2-1}$$

We have that $\{X, Y\}$ is an orthonormal basis of the horizontal distribution $\mathcal{H}$ and T is the Reeb vector field. The scalar product of two vector fields W and V with respect to the metric induced by the basis $\{X, Y, T\}$ will be often denoted by $\langle W, V\rangle$. This structure of $E(1, 1)$ is characterized by the following Lie brackets, [Milnor 1976],

$$[X, Y] = -T, \quad [X, T] = -Y, \quad [Y, T] = 0. \tag{2-2}$$

In fact, applying [Galli 2013, (9.1) and (9.3)] we obtain that the Webster scalar curvature is $W = -\frac{1}{2}$ and the matrix of the pseudohermitian torsion τ in the X, Y, T

basis is

$$\begin{pmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The following derivatives can be easily computed:

$$\begin{aligned} &\nabla_X X = 0, \quad \nabla_Y X = 0, \quad \nabla_T X = \tfrac{1}{2} Y, \\ &\nabla_X Y = 0, \quad \nabla_Y Y = 0, \quad \nabla_T Y = -\tfrac{1}{2} X, \end{aligned} \tag{2-3}$$

where ∇ denotes the pseudohermitian connection; see [Dragomir and Tomassini 2006]. Moreover we have $-2|\tau|^2 = W < 0$, which characterizes $E(1, 1)$; see [Perrone 1998]. We also define the involution J on $\mathcal{H}$, called the complex structure, by $J(X) = Y$ and $J(Y) = -X$.

The geometry of regular surfaces in $E(1,1)$. Consider a C^1 surface Σ immersed in $E(1, 1)$. We define the *subriemannian area* of Σ as

$$A(\Sigma) = \int_\Sigma |N_h| \, d\Sigma,$$

where N_h denotes the projection of the Riemannian unit normal N to $\mathcal{H}$ and $d\Sigma$ denotes the Riemannian area element on Σ. In the sequel we always denote by N the inner unit normal. The singular set Σ_0 is composed of the points in which $T\Sigma$ coincides with $\mathcal{H}$. Outside Σ_0, we can define the *horizontal unit normal* as

$$\nu_h := \frac{N_h}{|N_h|}$$

and the *characteristic vector field* as $Z := J(\nu_h)$. It is straightforward to verify that $\{Z, S\}$ is an orthonormal basis of $T\Sigma$ outside Σ_0, where

$$S := \langle N, T\rangle \nu_h - |N_h| T.$$

Finally, outside Σ_0, we define the *mean curvature* of Σ by

$$H := -\langle \nabla_Z \nu_h, Z\rangle. \tag{2-4}$$

Given a surface Σ as the zero level set of a function $u : \Omega \subset E(1, 1) \to \mathbb{R}$, we can express

$$\nu_h = -\frac{u_z X + \frac{1}{\sqrt{2}}(-e^z u_x + e^{-z} u_y) Y}{\sqrt{u_z^2 + \frac{1}{2}(-e^z u_x + e^{-z} u_y)^2}} \tag{2-5}$$

and

$$Z = \frac{\frac{1}{\sqrt{2}}(-e^z u_x + e^{-z} u_y) X - u_z Y}{\sqrt{u_z^2 + \frac{1}{2}(-e^z u_x + e^{-z} u_y)^2}}. \tag{2-6}$$

We define a *minimal surface* as a surface with vanishing mean curvature H.

Proposition 2.1. *Let Σ be a minimal surface defined as the zero level set of a C^2 function $u : \Omega \subset E(1,1) \to \mathbb{R}$. Then u satisfies the equation*

$$\begin{aligned} u_{zz}(-e^z u_x + e^{-z} u_y)^2 + u_z^2(-e^{2z} u_{xx} - 2u_{xy} + e^{-2z} u_{yy}) \\ - u_z(-e^z u_x + e^{-z} u_y)(-2e^z u_{xz} - e^z u_x + 2e^{-z} u_{yz} - e^{-z} u_y) = 0 \end{aligned} \tag{2-7}$$

on Ω.

Proof. From (2-4), (2-5) and (2-6) we find that u has to satisfy

$$Y(u)^2 X(X(u)) - Y(u)X(u)Y(X(u)) - Y(u)X(u)X(Y(u)) + X(u)^2 Y(Y(u)) = 0 \tag{2-8}$$

on Ω. Now, using (2-1), we can transform (2-8) into (2-7). □

We will call (2-7) the *minimal surface equation*.

Remark 2.2. From (2-8), we immediately note that a surface Σ satisfying $u_z \equiv 0$ or $-e^z u_x + e^{-z} u_y \equiv 0$ is always minimal.

In the following lemma, we compute some important quantities related to the torsion and the geometry of a surface. The lemma follows from [Galli 2013, (9.8)].

Lemma 2.3. *Let Σ be a C^1 surface in $E(1,1)$. Then we have*

$$\begin{aligned} &\langle \tau(Z), Z \rangle = -\langle Z, X \rangle \langle Z, Y \rangle = \langle \nu_h, X \rangle \langle \nu_h, Y \rangle = -\langle \tau(\nu_h), \nu_h \rangle, \\ &\langle \tau(Z), \nu_h \rangle = \tfrac{1}{2}(\langle Z, Y \rangle^2 - \langle Z, X \rangle^2). \end{aligned}$$

3. Characteristic curves in $E(1,1)$

In this section we will study the equation of the integral curves of Z on Σ, known as *characteristic curves*. It is well-known that a surface with constant mean curvature H is foliated by characteristic curves in $\Sigma - \Sigma_0$. In general, a *characteristic curve* is an arc-length parametrized horizontal curve γ in $E(1,1)$ that satisfies the equation

$$\nabla_{\dot\gamma} \dot\gamma + H J(\dot\gamma) = 0, \tag{3-1}$$

where $\dot\gamma$ denotes the tangent vector along γ and H is the (constant) curvature of γ. We stress that a curve γ satisfying (3-1) is not a subriemannian geodesic. In fact a characteristic curve γ is a subriemannian geodesic if and only if $H = 0$ and $\dot\gamma$ satisfies the additional equation

$$\langle \tau(\dot\gamma), \dot\gamma \rangle = 0, \tag{3-2}$$

see [Rumin 1994, Proposition 15], which forces γ to be an integral curve of X or Y, by Lemma 2.3.

Proposition 3.1. *Let γ be a characteristic curve in $E(1,1)$ with curvature $H=0$. Then γ belongs to the family of curves*

$$\gamma(t)=(x_0+\dot{x}_0t,\ y_0+\dot{y}_0t,\ z_0) \tag{3-3}$$

or to the family

$$\gamma(t)=\Big(x_0+\frac{\dot{x}_0}{\dot{z}_0}(e^{\dot{z}_0t}-1),\ y_0-\frac{\dot{y}_0}{\dot{z}_0}(e^{-\dot{z}_0t}-1),\ z_0+\dot{z}_0t\Big), \tag{3-4}$$

where $\gamma(0)=(x_0,y_0,z_0)$ and $\dot{\gamma}(0)=(\dot{x}_0,\dot{y}_0,\dot{z}_0)$.

Proof. We consider the curve $\gamma: I\to\Sigma$, where I denotes an interval. We express $\gamma(t)=(x(t),y(t),z(t))$ and we get

$$\begin{aligned}\dot{\gamma}(t)&=\dot{x}\frac{\partial}{\partial x}+\dot{y}\frac{\partial}{\partial y}+\dot{z}\frac{\partial}{\partial z}\\&=\dot{z}X+\frac{1}{\sqrt{2}}(\dot{y}e^{z}-\dot{x}e^{-z})Y+\frac{1}{\sqrt{2}}(\dot{y}e^{z}+\dot{x}e^{-z})T,\end{aligned} \tag{3-5}$$

since

$$\frac{\partial}{\partial x}=\frac{1}{\sqrt{2}}e^{-z}(T-Y),\qquad \frac{\partial}{\partial y}=\frac{1}{\sqrt{2}}e^{z}(Y+T).$$

From (3-5) and the fact that γ is horizontal, we have

$$\dot{y}e^{z}+\dot{x}e^{-z}=0. \tag{3-6}$$

Now $\nabla_{\dot{\gamma}}\dot{\gamma}=0$ is equivalent to the system

$$\begin{cases}\dot{z}=\dot{z}_0,\\ \dot{y}e^{z}-\dot{x}e^{-z}=c_0,\end{cases} \tag{3-7}$$

where $\dot{z}_0$ and c_0 are constants. We distinguish two cases. The first one corresponds to $\dot{z}_0=0$. This means which $z=z_0$, with $z_0\in\mathbb{R}$, and so (3-6) and (3-7) reduce to

$$\begin{cases}2\dot{y}=e^{-z_0}c_0,\\ 2\dot{x}=-e^{z_0}c_0,\end{cases} \tag{3-8}$$

which implies $\gamma(t)=(x_0-e^{z_0}(c_0/2)t,\ y_0+e^{-z_0}(c_0/2)t,\ z_0)$, where $c_0\neq 0$ and $x_0, y_0\in\mathbb{R}$.

The second possibility is $\dot{z}_0\neq 0$, which implies $z(t)=z_0+\dot{z}_0t$, with $z_0\in\mathbb{R}$. In this case integrating (3-8) we obtain

$$\gamma(t)=\Big(x_0+\frac{c_0e^{z_0}}{2\dot{z}_0}-\frac{c_0}{2\dot{z}_0}e^{z_0+\dot{z}_0t},\ y_0+\frac{c_0e^{-z_0}}{2\dot{z}_0}-\frac{c_0}{2\dot{z}_0}e^{-(z_0+\dot{z}_0)t},\ z_0+\dot{z}_0t\Big),$$

where $\gamma(0)=(x_0,y_0,z_0)$. Finally, to conclude the result, we note that

$$\frac{c_0}{2}=\dot{y}_0e^{z_0}=-\dot{x}_0e^{-z_0}. \qquad \square$$

4. Complete area-stationary surfaces with nonempty singular set in $E(1, 1)$

Complete area-stationary surfaces containing isolated singular points. The local structure of a C^1 surface Σ with prescribed mean curvature $H \in C$, in a neighborhood of an isolated singular point, is well understood [Cheng et al. 2012, Theorem D and Corollary E]. In the less general case of a bounded mean curvature surface of class C^2, applying [Cheng et al. 2005, Theorem B and Section 7], we have:

Lemma 4.1. *Let Σ be a C^2 oriented immersed surface with constant mean curvature H in $E(1, 1)$. If $p \in \Sigma_0$ is an isolated singular point, then there exists $r > 0$ and $\lambda \in \mathbb{R}$ such that the set*

$$D_r(p) = \{\gamma^H_{p,v}(s) : v \in T_p\Sigma,\ |v| = 1,\ s \in [0, r)\},$$

is an open neighborhood of p in Σ, where $\gamma^H_{p,v}$ denotes the characteristic curve starting from p in the direction v with curvature H.

First we construct the unique example, up to contact isometries, of a minimal surface with isolated singular points.

Proposition 4.2. *Let Σ be a C^2 complete, area-stationary surface immersed in $E(1, 1)$ with $H = 0$ and with an isolated singular point $p_0 = (x_0, y_0, z_0)$. Then $\Sigma = \{(x, y, z) \in E(1, 1) : e^{z-z_0}(y - y_0) + x - x_0 = 0\}$.*

Proof. By Lemma 4.1, the only possible way to construct a complete area-stationary surface, with a singular point p_0, is to consider the union of all characteristic curves γ of curvature 0 with initial conditions $\gamma(0) = p_0$ and $\dot\gamma(0) \in T_{p_0}\Sigma = \mathcal{H}_{p_0}$, $|\dot\gamma(0)| = 1$. We can suppose $p_0 = 0$, since $E(1, 1)$ is homogeneous.

We consider the initial velocities

$$\begin{aligned}\dot\gamma_\alpha(0) &= \cos\alpha\, X(0) + \sin\alpha\, Y(0)\\ &= \cos\alpha\, \frac{\partial}{\partial z}(0) + \frac{\sin\alpha}{\sqrt{2}}\Big(-\frac{\partial}{\partial x}(0) + \frac{\partial}{\partial y}(0)\Big),\end{aligned}$$

for $\alpha \in [0, 2\pi[$. In this way we obtain as characteristic curves

$$\begin{aligned}\gamma_\alpha(t)\\ = \Big(-\frac{\sin\alpha}{\sqrt{2}\cos\alpha}(e^{\cos(\alpha)t} - 1),\ -\frac{\sin\alpha}{\sqrt{2}\cos\alpha}(e^{-\cos(\alpha)t} - 1),\ \cos(\alpha)t\Big),\end{aligned} \tag{4-1}$$

for $\alpha \in \,]0, 2\pi[$ and $\gamma_0(t) = (0, 0, t)$ when $\alpha = 0$. At this point it is easy show that Σ is the zero level set of the function $e^z y + x$ (or equivalently $e^{-z}x + y$), which satisfies (2-7). $\square$

Complete area-stationary surfaces containing singular curves.

Lemma 4.3 [Galli 2013, Corollary 5.4]. *Let Σ be a C^2 minimal surface with nonempty singular set Σ_0 immersed in $E(1,1)$. Then Σ is area stationary if and only if the characteristic curves meet the singular curves orthogonally with respect the metric $\langle\ ,\ \rangle$, induced by the orthonormal basis* (2-1).

A minimal area-stationary surface cannot contain more than one singular curve:

Lemma 4.4. *Let Σ be a C^2 complete, minimal, area-stationary surface, containing a singular curve Γ immersed in $E(1,1)$. Then Σ cannot contain more singular curves.*

Proof. We consider a singular curve

$$\Gamma(\varepsilon) = \big(x(\varepsilon), y(\varepsilon), z(\varepsilon)\big)$$

in Σ. Since Σ is foliated by characteristic curves, we can parametrize it by the map

$$F(\varepsilon, t) = \gamma_\varepsilon(t) = \big(x(\varepsilon, t), y(\varepsilon, t), z(\varepsilon, t)\big),$$

where $\gamma_\varepsilon(t)$ is the characteristic curve with initial data $\gamma_\varepsilon(0) = \Gamma(\varepsilon)$ and

$$\begin{aligned} \dot\gamma_\varepsilon(0) &= J(\dot\Gamma(\varepsilon)) = \dot z(\varepsilon) J(X) + \tfrac{1}{\sqrt2}(\dot y(\varepsilon)e^{z(\varepsilon)} - \dot x(\varepsilon)e^{-z(\varepsilon)}) J(Y) \\ &= \tfrac{1}{\sqrt2}\big(-\dot z(\varepsilon)e^{z(\varepsilon)},\ \dot z(\varepsilon)e^{-z(\varepsilon)},\ \dot x(\varepsilon)e^{-z(\varepsilon)} - \dot y(\varepsilon)e^{z(\varepsilon)}\big). \end{aligned} \tag{4-2}$$

We define

$$V_\varepsilon(t) := \frac{\partial F}{\partial \varepsilon}(t, \varepsilon),$$

which is a smooth Jacobi-like vector field along $\gamma_\varepsilon(t)$; see [Galli 2013, Section 4]. At a singular point (ε, t), the vertical component of V_ε vanishes:

$$\langle V_\varepsilon, T\rangle(\varepsilon, t) = \frac{\partial x}{\partial \varepsilon}(\varepsilon, t)e^{-z(\varepsilon,t)} + \frac{\partial y}{\partial \varepsilon}(\varepsilon, t)e^{z(\varepsilon,t)} = 0.$$

We suppose that Γ is not an integral curve of X or Y. Then from the expressions of the component of $F(\varepsilon, t)$, which are

$$\begin{aligned} x(\varepsilon, t) &= x(\varepsilon) + \frac{\dot z(\varepsilon)e^{z(\varepsilon)}}{\dot x(\varepsilon)e^{-z(\varepsilon)} - \dot y(\varepsilon)e^{z(\varepsilon)}}\big(e^{(\dot x(\varepsilon)e^{-z(\varepsilon)} - \dot y(\varepsilon)e^{z(\varepsilon)})t/\sqrt2} - 1\big), \\ y(\varepsilon, t) &= y(\varepsilon) - \frac{\dot z(\varepsilon)e^{-z(\varepsilon)}}{\dot x(\varepsilon)e^{-z(\varepsilon)} - \dot y(\varepsilon)e^{z(\varepsilon)}}\big(e^{-(\dot x(\varepsilon)e^{-z(\varepsilon)} - \dot y(\varepsilon)e^{z(\varepsilon)})t/\sqrt2} - 1\big), \\ z(\varepsilon, t) &= z(\varepsilon) + \frac{\dot x(\varepsilon)e^{-z(\varepsilon)} - \dot y(\varepsilon)e^{z(\varepsilon)}}{\sqrt2} t, \end{aligned} \tag{4-3}$$

we have

$$
\begin{aligned}
&\langle V_\varepsilon, T\rangle(\varepsilon, t)\\
&=\left(\dot{x}(\varepsilon)e^{-z(\varepsilon)}+\frac{\ddot{z}(\varepsilon)}{\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)}}-\frac{\dot{z}(\varepsilon)\dfrac{\partial}{\partial\varepsilon}(\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)})}{(\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)})^2}\right)\\
&\qquad\qquad\cdot\left(e^{-(\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)})\frac{t}{\sqrt{2}}}-e^{(\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)})\frac{t}{\sqrt{2}}}\right)\\
&\quad+\frac{\dot{z}(\varepsilon)^2}{\left(\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)}\right)^2}\left(e^{-(\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)})\frac{t}{\sqrt{2}}}+e^{(\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)})\frac{t}{\sqrt{2}}}-2\right),
\end{aligned}
$$

which, when t is positive, vanishes only for the values $(\varepsilon, 0)$. On the other hand, if Γ is an integral curve of Y we get

$$
x(\varepsilon,t)=x(\varepsilon),\quad y(\varepsilon,t)=y(\varepsilon),\quad z(\varepsilon,t)=z(\varepsilon)+\frac{\dot{x}(\varepsilon)e^{-z(\varepsilon)}-\dot{y}(\varepsilon)e^{z(\varepsilon)}}{\sqrt{2}}t, \tag{4-4}
$$

and if Γ is an integral curve of X we have

$$
x(\varepsilon,t)=x(\varepsilon)-\frac{\dot{z}(\varepsilon)e^{z(\varepsilon)}}{\sqrt{2}}t,\quad y(\varepsilon,t)=y(\varepsilon)+\frac{\dot{z}(\varepsilon)e^{-z(\varepsilon)}}{\sqrt{2}}t,\quad z(\varepsilon,t)=z(\varepsilon). \tag{4-5}
$$

In both cases, the singular set is only the curve $\Gamma(\varepsilon)$. □

The vertical component of V_ε can be computed more directly using [Galli 2013, Proposition 4.3], since $H=0$. On the other hand, the explicit computation of the components of the parametrization $F(\varepsilon,t)$ allows us to characterize all C^2 area-stationary complete surfaces with a singular curve that is a characteristic curve of curvature 0. We stress that, when the characteristic curves are subriemannian geodesics, these examples can also be constructed from Remark 2.2.

Proposition 4.5. *Let Σ be an area-stationary surface with $H=0$, with a singular curve Γ that is a characteristic curve of curvature 0. Then, if Γ is a subriemannian geodesic, Σ belongs to one of the following families*:

(i) $\{ax+by+c=0:(x,y,z)\in E(1,1),\, a,b,c\in\mathbb{R}\}$;

(ii) $\{e^{z-z_0}(y-y_0)+e^{z_0-z}(x-x_0)=0:(x,y,z)\in E(1,1),\, x_0,y_0,z_0\in\mathbb{R}\}$.

Otherwise, we suppose that Γ is a characteristic curve passing through (x_0,y_0,z_0) with velocity $(\dot{x}_0,\dot{y}_0,\dot{z}_0)$, $\dot{x}_0,\dot{y}_0,\dot{z}_0\neq 0$. We can parametrize Σ by $F:\mathbb{R}^2\to E(1,1)$, with $F(\varepsilon,t)=(x(\varepsilon,t),y(\varepsilon,t),z(\varepsilon,t))$ and

$$
\begin{aligned}
x(\varepsilon,t) &= x_0 + \frac{\dot{x}_0}{\dot{z}_0}(e^{\dot{z}_0\varepsilon}-1) + \frac{\dot{z}_0 e^{z_0+\dot{z}_0\varepsilon}}{\dot{x}_0 e^{-z_0}-\dot{y}_0 e^{z_0}}(e^{(\dot{x}_0 e^{-z_0}-\dot{y}_0 e^{z_0})t/\sqrt{2}}-1),\\
y(\varepsilon,t) &= y_0 - \frac{\dot{y}_0}{\dot{z}_0}(e^{-\dot{z}_0\varepsilon}-1) - \frac{\dot{z}_0 e^{-z_0-\dot{z}\varepsilon}}{\dot{x}_0 e^{-z_0}-\dot{y}_0 e^{z_0}}(e^{-(\dot{x}_0 e^{-z_0}-\dot{y}_0 e^{z_0})t/\sqrt{2}}-1),\\
z(\varepsilon,t) &= z_0 + \dot{z}_0\varepsilon + \frac{\dot{x}_0 e^{-z_0}-\dot{y}_0 e^{z_0}}{\sqrt{2}}t.
\end{aligned}
\tag{4-6}
$$

Remark 4.6. The surfaces parametrized by (4-6) are the first examples of area-stationary surfaces that are not foliated by subriemannian geodesics in three-dimensional contact subriemannian manifolds, up to our knowledge. In fact this phenomenon does not appear in the group of rigid motions [Galli 2013, Lemma 10.4], even if its pseudohermitian torsion is nonvanishing. In that case, the presence of two singular curves forces the surface to be foliated by subriemannian geodesics or to be not area-stationary. On the other hand, it is well-known that a minimal surface is foliated by subriemannian geodesics in any three-dimensional Sasakian manifold.

Remark 4.7. Given any horizontal curve $\Gamma = (x(\varepsilon), y(\varepsilon), z(\varepsilon))$ in $E(1,1)$, we stress that (4-3) provides a parametrization $F(\varepsilon,t) : \mathbb{R}^2 \to \Sigma \subset E(1,1)$ of a complete area-stationary surface Σ with $\Sigma_0 = \Gamma$.

5. Complete area-minimizing surfaces in $E(1,1)$

Complete area-minimizing surfaces with empty singular set. Proposition 9.8 of [Galli 2013] gave a general necessary condition for the stability of a nonsingular surface in pseudohermitian Lie groups. This condition states that the quantity

$$
W - \langle \tau(Z), \nu_h\rangle = \langle \nu_h, Y\rangle^2 - 1 = \langle Z, X\rangle^2 - 1
$$

must be always nonpositive. This condition is trivial in $E(1,1)$ due to the negativity of the Webster scalar curvature. On the other hand it has been used crucially in the classification of the stable, area-stationary surfaces without singular points in the manifolds $\mathbb{H}^1$, SU(2) and $\widetilde{E(2)}$, see [Galli 2013; Hurtado et al. 2010; Rosales 2012]. In any case, we can prove:

Proposition 5.1. *The families of planes*

(i) $\{x + c = 0 : (x,y,z) \in E(1,1), c \in \mathbb{R}\}$,

(ii) $\{y + c = 0 : (x,y,z) \in E(1,1), c \in \mathbb{R}\}$,

(iii) $\{z + c = 0 : (x,y,z) \in E(1,1), c \in \mathbb{R}\}$,

are area-stationary, foliated by subriemannian geodesics, and area-minimizing.

Proof. We prove the result for $\Sigma = \{x = 0 : (x,y,z) \in E(1,1)\}$, since all the cases are similar. In this case, from (2-5) and (2-6) we have

$$
\nu_h = Y, \quad Z = -X.
$$

So the integral curves of Z are subriemannian geodesics and $\Sigma_0 = \varnothing$. Now Remark 2.2 implies that Σ is area-stationary. Finally we can foliate a neighborhood of Σ in $E(1, 1)$ by translating Σ. We obtain a foliation by area-stationary surfaces, and a standard calibration argument implies that Σ is area-minimizing; see, for example, [Barone Adesi et al. 2007; Ritoré 2009; Ritoré and Rosales 2008, § 5]. □

Remark 5.2. The planes in the family

$$\{ax + by + cz + d = 0 : (x, y, z) \in E(1, 1), a, b, c, d \in \mathbb{R}\}$$

are not minimal, since they do not satisfy (2-7).

A very natural question is: are the planes in Proposition 5.1 the unique complete area-minimizing surfaces with empty singular set in $E(1, 1)$? We have only been able to find the following sufficient condition:

Lemma 5.3. *Let Σ be a C^2 complete oriented minimal surface immersed in $E(1, 1)$, with empty singular set Σ_0. If $\langle N, T\rangle \leqslant 0$ holds on Σ, then Σ is stable.*

Proof. Taking into account the expression of the stability operator for nonsingular surfaces in [Galli 2013, Lemma 8.3], we only need to show that

$$2Z(G) + G^2 \leqslant 0 \quad \text{on } \Sigma, \qquad \text{where } G := \frac{\langle N, T\rangle}{|N_h|}.$$

Given a point p in Σ, let I be an open interval containing the origin and let $\alpha : I \to \Sigma$ be a piece of the integral curve of S passing through p. Consider the characteristic curve $\gamma_\varepsilon(s)$ of Σ with $\gamma_\varepsilon(0) = \alpha(\varepsilon)$. We define the map $F : I \times \mathbb{R} \to \Sigma$ by $F(\varepsilon, s) = \gamma_\varepsilon(s)$ and set $V(s) := (\partial F/\partial\varepsilon)(0, s)$, which is a Jacobi-like vector field along γ_0; see [ibid., Proposition 4.3]. Let $'$ represent differentiation with respect to s. Using [ibid., Lemma 3.1, (4.4) and (4.5)] we get

$$\langle V, T\rangle(0) = -|N_h|, \tag{5-1}$$

$$\langle V, T\rangle'(0) = -\langle N, T\rangle, \tag{5-2}$$

$$\langle V, T\rangle''(0) = -|N_h|(Z(G) + G^2). \tag{5-3}$$

It is easy to show that $g(V, T)$ never vanishes along γ_0 since Σ_0 is empty; see [ibid., proof of Proposition 9.5]. On the other hand, by [ibid., Proposition 4.3] and Lemma 2.3, we have that $\langle V, T\rangle$ satisfies the ordinary differential equation

$$\langle V, T\rangle'''(s) - \langle Z, X\rangle^2 \langle V, T\rangle'(s) = 0$$

along γ_0. We suppose that $\langle Z, X\rangle \neq 0$. Taking into account the initial conditions (5-1), (5-2) and (5-3), we obtain

$$\langle V, T\rangle(s) = a \cosh(|\langle Z, X\rangle| s) + b \sinh(|\langle Z, X\rangle| s) + c,$$

where

$$a = \frac{|N_h|(Z(G)+G^2)}{\langle X, Z\rangle^2}, \quad b = -\frac{\langle N, T\rangle}{|\langle Z, X\rangle|}, \quad c = -|N_h| - a.$$

We have that $\langle V, T\rangle(s) \neq 0$ implies

$$a + b = \frac{|N_h|(Z(G)+G^2)}{\langle X, Z\rangle^2} - \frac{\langle N, T\rangle}{|\langle Z, X\rangle|} \leqslant 0.$$

Then we can conclude that

$$2Z(G) + G^2 \leqslant 2(Z(G)+G^2) \leqslant 2|\langle Z, X\rangle| \frac{\langle N, T\rangle}{|N_h|} \leqslant 0$$

on γ_0. Now since the choice of p is arbitrary, we get the statement.

If $\langle Z, X\rangle = 0$, we conclude that Σ is stable if and only if $\langle N, T\rangle = 0$, by [Galli 2013, Proposition 9.8]. □

Remark 5.4. The surfaces described in the points (i), (ii) and (iii) of Proposition 5.1 are characterized by $\langle N, T\rangle = -e^z/\sqrt{2}$, $\langle N, T\rangle = -e^z/\sqrt{2}$ and $\langle N, T\rangle \equiv 0$, respectively, where N denotes the inward unit normal on Σ. In the third family the planes are vertical surfaces and they satisfy $W - \langle \tau(Z), \nu_h\rangle \equiv 0$.

Taking into account the geometric invariants of $E(1, 1)$, we expect the existence of other examples of complete oriented minimal surface with empty singular set.

Complete area-minimizing surfaces with nonempty singular set. We consider the stability operator constructed in [Galli 2013, Theorem 8.6].

Lemma 5.5. *Let Σ be a C^2 oriented minimal surface immersed in E(1,1), with singular set Σ_0 and $\partial\Sigma = \varnothing$. If Σ is stable then, for any function $u \in C_0^1(\Sigma)$ such that $Z(u) = 0$ in a tubular neighborhood of a singular curve and constant in a tubular neighborhood of an isolated singular point, we have $Q(u) \geqslant 0$, where*

$$Q(u) := \int_\Sigma \Big\{ |N_h|^{-1} Z(u)^2 + |N_h| \big((1+\langle Z, Y\rangle^2) - \big(|N_h|(\tfrac{1}{2} - \langle Z, Y\rangle^2) - \langle \nabla_S \nu_h, Z\rangle \big)^2 \big) u^2 \Big\}\, d\Sigma$$
$$+ 4 \int_{(\Sigma_0)_c} \langle N, T\rangle \langle Z, Y\rangle^2 \langle Z, \nu\rangle u^2 \, d(\Sigma_0)_c + \int_{(\Sigma_0)_c} S(u)^2 \, d(\Sigma_0)_c.$$

Here $d(\Sigma_0)_c$ is the Riemannian length measure on $(\Sigma_0)_c$ and ν is the external unit normal to $(\Sigma_0)_c$.

Corollary 5.6. *Let Σ be a plane in the family*

$$\{ax + by + c = 0 : (x, y, z) \in E(1, 1), a, b, c \in \mathbb{R}\}.$$

Then Σ is stable.

Proof. We know that Σ is area-stationary with a singular line, obtained intersecting Σ with the plane $z = \log\sqrt{b/a}$. From (2-6) we get

$$Z = \frac{-be^{z} + ae^{-z}}{|-be^{z} + ae^{-z}|} X,$$

which is orthogonal to the singular line. Since $\langle \nabla_S \nu_h, Z\rangle = \langle \nabla_S Y, X\rangle = \dfrac{|N_h|}{2}$, the stability operator

$$Q(u) = \int_{\Sigma} \{|N_h|^{-1} Z(u)^2 + |N_h| \langle N, T\rangle^2 u^2\}\, d\Sigma + \int_{\Sigma_0} S(u)^2 \, d\Sigma_0$$

is always nonnegative for any admissible test function u. □

Remark 5.7. The planes $\{ax + by + c = 0 : (x, y, z) \in E(1, 1), a, b, c \in \mathbb{R}\}$ are also area-minimizing, by calibration arguments.

Corollary 5.8. *The surface* $\Sigma = \{e^{z} y + e^{-z} x = 0 : (x, y, z) \in E(1, 1)\}$ *is stable.*

Proof. From (2-6) we get

$$Z = -\frac{(e^{z} y - e^{-z} x) Y}{|e^{z} y - e^{-z} x|}$$

and $\Sigma_0 = \{(0, 0, z) : (x, y, z) \in E(1, 1)\}$. From (2-3) we have

$$\langle \nabla_S \nu_h, Z\rangle = \langle \nabla_S Y, X\rangle = -\frac{|N_h|}{2},$$

which implies

$$Q(u) = \int_{\Sigma} \{|N_h|^{-1} Z(u)^2 + 2|N_h|^2 u^2\}\, d\Sigma + \int_{\Sigma_0} S(u)^2 \, d\Sigma_0 + 4 \int_{\Sigma_0} u^2 \, d\Sigma_0 \geqslant 0,$$

for all admissible u. □

Corollary 5.9. *The surfaces defined in Proposition 4.2 are stable.*

Proof. For simplicity we will prove the statement in the case of $x_0 = y_0 = z_0 = 0$. We note that, since $\Sigma_0 = (0, 0, 0)$, the argument in the proof of Lemma 5.3 works and the condition $\langle N, T\rangle = -(1 + e^{z})/\sqrt{2} \leqslant 0$ is sufficient for the stability in the complement of any tubular neighborhood of Σ_0. Finally we observe that the stability operator in Lemma 5.5 makes no contribution to the singular set in the case of isolated singular points. □

References

[Ambrosio et al. 2006] L. Ambrosio, F. Serra Cassano, and D. Vittone, "Intrinsic regular hypersurfaces in Heisenberg groups", *J. Geom. Anal.* **16**:2 (2006), 187–232. MR 2007g:49072 Zbl 1085.49045

[Barbieri and Citti 2011] D. Barbieri and G. Citti, "Regularity of minimal intrinsic graphs in 3-dimensional sub-Riemannian structures of step 2", *J. Math. Pures Appl.* (9) **96**:3 (2011), 279–306. MR 2012g:53050 Zbl 1231.53030

[Barone Adesi et al. 2007] V. Barone Adesi, F. Serra Cassano, and D. Vittone, "The Bernstein problem for intrinsic graphs in Heisenberg groups and calibrations", *Calc. Var. Partial Differential Equations* **30**:1 (2007), 17–49. MR 2009c:35044 Zbl 1206.35240

[Bigolin and Cassano 2010] F. Bigolin and F. S. Cassano, "Distributional solutions of Burgers' equation and intrinsic regular graphs in Heisenberg groups", *J. Math. Anal. Appl.* **366**:2 (2010), 561–568. MR 2012a:35335 Zbl 1186.35030

[Boothby and Wang 1958] W. M. Boothby and H. C. Wang, "On contact manifolds", *Ann. of Math.* (2) **68** (1958), 721–734. MR 22 #3015 Zbl 0084.39204

[Capogna et al. 2007] L. Capogna, D. Danielli, S. D. Pauls, and J. T. Tyson, *An introduction to the Heisenberg group and the sub-Riemannian isoperimetric problem*, Progress in Mathematics **259**, Birkhäuser, Basel, 2007. MR 2009a:53053 Zbl 1138.53003

[Capogna et al. 2009] L. Capogna, G. Citti, and M. Manfredini, "Regularity of non-characteristic minimal graphs in the Heisenberg group $\mathbb{H}^1$", *Indiana Univ. Math. J.* **58**:5 (2009), 2115–2160. MR 2010j:58032 Zbl 1182.35087

[Cheng and Hwang 2010] J.-H. Cheng and J.-F. Hwang, "Variations of generalized area functionals and p-area minimizers of bounded variation in the Heisenberg group", *Bull. Inst. Math. Acad. Sin. (N.S.)* **5**:4 (2010), 369–412. MR 2012g:49085 Zbl 1232.35178

[Cheng et al. 2005] J.-H. Cheng, J.-F. Hwang, A. Malchiodi, and P. Yang, "Minimal surfaces in pseudohermitian geometry", *Ann. Sc. Norm. Super. Pisa Cl. Sci.* (5) **4**:1 (2005), 129–177. MR 2006f:53008 Zbl 1158.53306

[Cheng et al. 2007] J.-H. Cheng, J.-F. Hwang, and P. Yang, "Existence and uniqueness for p-area minimizers in the Heisenberg group", *Math. Ann.* **337**:2 (2007), 253–293. MR 2009h:35120 Zbl 1109.35009

[Cheng et al. 2012] J.-H. Cheng, J.-F. Hwang, A. Malchiodi, and P. Yang, "A Codazzi-like equation and the singular set for C^1 smooth surfaces in the Heisenberg group", *J. Reine Angew. Math.* **671** (2012), 131–198. MR 2983199 Zbl 1255.53026

[Danielli et al. 2007] D. Danielli, N. Garofalo, and D. M. Nhieu, "Sub-Riemannian calculus on hypersurfaces in Carnot groups", *Adv. Math.* **215**:1 (2007), 292–378. MR 2009h:53061 Zbl 1129.53017

[Danielli et al. 2008] D. Danielli, N. Garofalo, and D. M. Nhieu, "A notable family of entire intrinsic minimal graphs in the Heisenberg group which are not perimeter minimizing", *Amer. J. Math.* **130**:2 (2008), 317–339. MR 2009b:49102 Zbl 1158.53334

[Danielli et al. 2009] D. Danielli, N. Garofalo, D. M. Nhieu, and S. D. Pauls, "Instability of graphical strips and a positive answer to the Bernstein problem in the Heisenberg group $\mathbb{H}^1$", *J. Differential Geom.* **81**:2 (2009), 251–295. MR 2010e:53007 Zbl 1161.53024

[Dragomir and Tomassini 2006] S. Dragomir and G. Tomassini, *Differential geometry and analysis on CR manifolds*, Progress in Mathematics **246**, Birkhäuser, Boston, MA, 2006. MR 2007b:32056 Zbl 1099.32008

[Galli 2012] M. Galli, *Area-stationary surfaces in contact sub-Riemannian manifolds*, Ph.D. thesis, Universidad de Granada, 2012, Available at http://hera.ugr.es/tesisugr/21013020.pdf.

[Galli 2013] M. Galli, "First and second variation formulae for the sub-Riemannian area in three-dimensional pseudo-Hermitian manifolds", *Calc. Var. Partial Differential Equations* **47**:1-2 (2013), 117–157. MR 3044134 Zbl 1268.53039

[Galli and Ritoré 2013] M. Galli and M. Ritoré, "Existence of isoperimetric regions in contact sub-Riemannian manifolds", *J. Math. Anal. Appl.* **397**:2 (2013), 697–714. MR 2979606 Zbl 06109993

[Garofalo and Nhieu 1996] N. Garofalo and D.-M. Nhieu, "Isoperimetric and Sobolev inequalities for Carnot–Carathéodory spaces and the existence of minimal surfaces", *Comm. Pure Appl. Math.* **49**:10 (1996), 1081–1144. MR 97i:58032 Zbl 0880.35032

[Hladky and Pauls 2008] R. K. Hladky and S. D. Pauls, "Constant mean curvature surfaces in sub-Riemannian geometry", *J. Differential Geom.* **79**:1 (2008), 111–139. MR 2009m:53070 Zbl 1156.53038

[Hurtado and Rosales 2008] A. Hurtado and C. Rosales, "Area-stationary surfaces inside the sub-Riemannian three-sphere", *Math. Ann.* **340**:3 (2008), 675–708. MR 2008i:53038 Zbl 1132.53015

[Hurtado and Rosales $\geq$ 2014] A. Hurtado and C. Rosales, "Stable surfaces inside the sub-Riemannian three-sphere". In preparation.

[Hurtado et al. 2010] A. Hurtado, M. Ritoré, and C. Rosales, "The classification of complete stable area-stationary surfaces in the Heisenberg group $\mathbb{H}^1$", *Adv. Math.* **224**:2 (2010), 561–600. MR 2011d:53054 Zbl 1192.53038

[Milnor 1976] J. Milnor, "Curvatures of left invariant metrics on Lie groups", *Advances in Math.* **21**:3 (1976), 293–329. MR 54 #12970 Zbl 0341.53030

[Perrone 1998] D. Perrone, "Homogeneous contact Riemannian three-manifolds", *Illinois J. Math.* **42**:2 (1998), 243–256. MR 99a:53067 Zbl 0906.53031

[Ritoré 2009] M. Ritoré, "Examples of area-minimizing surfaces in the sub-Riemannian Heisenberg group $\mathbb{H}^1$ with low regularity", *Calc. Var. Partial Differential Equations* **34**:2 (2009), 179–192. MR 2009h:53062 Zbl 1165.53023

[Ritoré and Rosales 2008] M. Ritoré and C. Rosales, "Area-stationary surfaces in the Heisenberg group $\mathbb{H}^1$", *Adv. Math.* **219**:2 (2008), 633–671. MR 2009h:49075 Zbl 1158.53022

[Rosales 2012] C. Rosales, "Complete stable CMC surfaces with empty singular set in Sasakian sub-Riemannian 3-manifolds", *Calc. Var. Partial Differential Equations* **43**:3-4 (2012), 311–345. MR 2875642 Zbl 1235.53036

[Rumin 1994] M. Rumin, "Formes différentielles sur les variétés de contact", *J. Differential Geom.* **39**:2 (1994), 281–330. MR 95g:58221 Zbl 0973.53524

[Shcherbakova 2009] N. Shcherbakova, "Minimal surfaces in sub-Riemannian manifolds and structure of their singular sets in the (2, 3) case", *ESAIM Control Optim. Calc. Var.* **15**:4 (2009), 839–862. MR 2011c:53055 Zbl 1190.53027

Received May 27, 2013. Revised August 9, 2013.

MATTEO GALLI
DEPARTAMENTO DE GEOMETRÍA Y TOPOLOGÍA
UNIVERSIDAD DE GRANADA
18071 GRANADA
SPAIN

galli@ugr.es

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.159

PERIODIC ORBITS OF HAMILTONIAN SYSTEMS LINEAR AND HYPERBOLIC AT INFINITY

BAŞAK Z. GÜREL

We consider Hamiltonian diffeomorphisms of symplectic Euclidean spaces, generated by compactly supported time-dependent perturbations of hyperbolic quadratic forms. We prove that, under some natural assumptions, such a diffeomorphism must have simple periodic orbits of arbitrarily large period when it has fixed points which are not necessary from a homological perspective.

1. Introduction and main results

Introduction. In this paper we consider time-dependent Hamiltonians H on $\mathbb{R}^{2n}$ which, outside a compact set, are autonomous and coincide with a hyperbolic quadratic form (i.e., a nondegenerate quadratic form whose Hamiltonian vector field has no purely imaginary eigenvalues). We prove that, under some additional conditions, the Hamiltonian diffeomorphism φ_H must have simple (i.e., uniterated) periodic orbits of arbitrarily large (prime) period when it has certain "homologically unnecessary" fixed points. In particular, φ_H then has infinitely many periodic orbits. To be more precise, this result holds provided that φ_H has at least one nondegenerate (or even homologically nontrivial) fixed point with nonzero mean index, and the quadratic form (i.e., the corresponding linear Hamiltonian vector field) has only real eigenvalues. (See Remark 3.3 for the case of complex eigenvalues.)

Our main motivation for studying this question comes from a variant of the Conley conjecture applicable to manifolds for which the standard Conley conjecture fails. Recall in this connection that the latter asserts the existence of infinitely

The work is partially supported by the NSF grants DMS-0906204 and DMS-1207680.
MSC2010: 37J10, 53D40.
Keywords: periodic orbits, Hamiltonian flows, Floer homology, Conley conjecture.

many periodic orbits for every Hamiltonian diffeomorphism of a closed symplectic manifold. This is the case for manifolds with spherically vanishing first Chern class (of the tangent bundle) and also for negative monotone manifolds; see [Chance et al. 2013; Ginzburg and Gürel 2009a; Hein 2012] and also [Franks and Handel 2003; Ginzburg 2010; Ginzburg and Gürel 2012; Hingston 2009; Le Calvez 2006; Salamon and Zehnder 1992]. However the Conley conjecture, as stated, fails for some simple manifolds, such as S^2: an irrational rotation of S^2 about the z-axis has only two periodic orbits, which are also the fixed points; these are the poles. In fact, any manifold that admits a Hamiltonian torus action with isolated fixed points also admits a Hamiltonian diffeomorphism with finitely many periodic orbits. In particular, $\mathbb{CP}^n$, the Grassmannians and, more generally, most of the coadjoint orbits of compact Lie groups as well as symplectic toric manifolds all admit Hamiltonian diffeomorphisms with finitely many periodic orbits.

A viable alternative to the Conley conjecture for such manifolds is the conjecture that a Hamiltonian diffeomorphism with more fixed points than necessarily required by the (weak) Arnold conjecture has infinitely many periodic orbits. (It is possible that in this conjecture one might need to impose some kind of nondegeneracy condition (e.g., homological nontriviality) on the fixed points, as is the case for the version considered in this paper.) For $\mathbb{CP}^n$, the expected threshold is $n+1$. This conjecture is inspired by a celebrated theorem of Franks [1992; 1996] stating that a Hamiltonian diffeomorphism (or even an area preserving homeomorphism) of S^2 with at least three fixed points must have infinitely many periodic orbits; see also [Franks and Handel 2003; Le Calvez 2006] for further refinements and [Bramham and Hofer 2012; Collier et al. 2012; Kerman 2012] for symplectic topological proofs. We will refer to this analogue of the Conley conjecture as the *HZ conjecture* since, to the best of our knowledge, the first written account of the assertion is in [Hofer and Zehnder 1994, p. 263].

We find it useful to view the HZ conjecture in a broader context. Namely it appears that the presence of a fixed point that is unnecessary from a homological or geometrical perspective is already sufficient to force the existence of infinitely many periodic orbits. For instance, a theorem from [Ginzburg and Gürel 2014] asserts that for a certain class of closed monotone symplectic manifolds including $\mathbb{CP}^n$, any Hamiltonian diffeomorphism with a hyperbolic fixed point must necessarily have infinitely many periodic orbits. (Note that the original HZ conjecture, at least for nondegenerate Hamiltonian diffeomorphisms of $\mathbb{CP}^n$, would follow if one could replace a hyperbolic fixed point with a nonelliptic one in this theorem.) Furthermore there are obvious analogues of the HZ conjecture for symplectomorphisms or noncontractible periodic orbits of Hamiltonian diffeomorphisms. These analogues are also of interest and in some instances more accessible than the original HZ conjecture; see, for example, [Batoreo 2013; Ginzburg and Gürel 2009b; 2014; Gürel 2013].

The generalized HZ conjecture is also the central theme of this paper, although here we focus on a different aspect of the problem. Our main result, Theorem 1.1, can be viewed as a "local version" of this conjecture, and it holds in all dimensions. Namely we prove a variant of the HZ conjecture for Hamiltonians on $\mathbb{R}^{2n}$ which are compactly supported perturbations of certain quadratic forms. Working with $\mathbb{R}^{2n}$ allows us to circumvent a number of symplectic topological obstacles to proving the HZ conjecture and concentrate on what we interpret as the dynamical part of the problem, which is still quite nontrivial. This is a key difference, technical and conceptual, between the present work and the approach taken in [Ginzburg and Gürel 2014], where the symplectic topology of the ambient manifold plays a central role. We use Floer-theoretical techniques in the proofs. Deferring a more detailed discussion of our method to Section 1, we merely mention at this point that for technical reasons the quadratic form needs to be hyperbolic. Finally it should also be noted that Hamiltonian systems on $\mathbb{R}^{2n}$ with a controlled (e.g., asymptotically linear) behavior at infinity have been extensively studied in the context of Hamiltonian mechanics by classical variational methods; see, for example, [Abbondandolo 2001; Amann and Zehnder 1980; Antonacci 1997; Cornea 2001; Mawhin and Willem 1989; Rabinowitz 1980; Zhang and Liu 2011; Zou 2001] and references therein. However, to the best of our knowledge, there is no overlap between that approach and the present work, including the results.

Main results. To state the main results of the paper, recall that the *mean index* $\Delta_H(x) \in \mathbb{R}$ of a periodic orbit x of the Hamiltonian flow of H measures, roughly speaking, the total angle swept out by certain eigenvalues with absolute value one of the linearized flow $d\varphi_H^t$ along x; see [Long 2002; Salamon and Zehnder 1992] and also [Entov and Polterovich 2009, Section 3.3] and references therein for a more detailed discussion. For instance, the mean index is zero when $d\varphi_H^t$ has no eigenvalues on the unit circle for any $t \neq 0$, and hence the orbit is hyperbolic. Finally denote by $\mathrm{Fix}(\varphi_H)$ the collection of fixed points of φ_H.

Theorem 1.1. *Let $H : S^1 \times \mathbb{R}^{2n} \to \mathbb{R}$ be a Hamiltonian which is equal to a hyperbolic quadratic form Q at infinity* (*i.e.*, *outside a compact set*) *such that Q has only real eigenvalues. Assume that φ_H has a nondegenerate fixed point with nonzero mean index and* $\mathrm{Fix}(\varphi_H)$ *is finite. Then φ_H has simple*, *that is*, *uniterated*, *periodic orbits of arbitrarily large period.*

As a consequence, φ_H has infinitely many simple periodic orbits regardless of whether $\mathrm{Fix}(\varphi_H)$ is finite or not. In fact, the nondegeneracy condition in Theorem 1.1 can be relaxed and replaced by a much weaker, albeit more technical, condition that the point is isolated and homologically nontrivial, that is, its local Floer homology is nonzero. This is Theorem 4.1.

Remark 1.2. Theorems 1.1 and 1.4 (below), and their generalizations discussed in Section 4, also hold when the quadratic form Q has complex eigenvalues σ, provided that $|\mathrm{Re}\,\sigma| > |\mathrm{Im}\,\sigma|$; see Remark 3.3.

Remark 1.3. From the perspective of the generalized HZ-conjecture, observe that the nondegenerate (or homologically nontrivial) fixed point with nonzero mean index in Theorem 1.1 is the "unnecessary" point. Moreover the presence of one such point x implies the existence of at least two other (homologically nontrivial) orbits. Indeed the Floer homology for all iterations of H is concentrated in degree zero (see Section 3), and once k is so large that the index of the iterated orbit x^k is outside the range $[-n, n]$, another orbit must take over generating the homology. Furthermore there should be at least one more periodic orbit to cancel out the contribution of x^k to the homology in higher degrees.

Hypothetically results similar to Theorem 1.1 and other theorems discussed in this section hold when a hyperbolic quadratic form is replaced by any (autonomous) quadratic form Q without nontrivial periodic orbits. For instance in this case, one can expect to have infinitely many periodic orbits whenever φ_H has a nondegenerate fixed point with mean index different from $\Delta_Q(0)$ or has at least two nondegenerate fixed points; cf. Remark 4.6. (The latter conjecture, which was the starting point of this work, is due to Alberto Abbondandolo.)

As has been pointed out above, the proof of Theorem 1.1 is based on Floer theory. However for a general quadratic form Q, even when the Floer homology exists, continuation maps fail to have the desired properties and the homology is not invariant under iterations. This is the case, for instance, for positive or negative definite Q (see Remark 3.6) and the main reason why we restrict our attention to hyperbolic quadratic forms. Even for such forms some foundational aspects of Floer theory have to be reexamined. We do this in Section 3, using, as one could expect, a version of the maximum principle.

The condition that the fixed point is nondegenerate (and that it has nonzero mean index) is essential in Theorem 1.1. For instance, starting with the flow of $Q(x, y) = xy$ on $\mathbb{R}^2$, it is easy to introduce degenerate (homologically trivial) fixed points by slightly perturbing the flow away from the saddle. This way one can create an arbitrarily large number of fixed points without generating infinitely many periodic orbits. In fact, we expect some form of nondegeneracy (e.g., homological nontriviality) to be essential in the HZ conjecture beyond the case of S^2.

In low dimensions, Theorem 1.1 combined with simple index analysis implies the HZ conjecture in its original form for Hamiltonians in question. To state the result, recall first that φ_H is said to be strongly nondegenerate if all iterations of φ_H are nondegenerate.

Theorem 1.4. *Let $H : S^1 \times \mathbb{R}^{2n} \to \mathbb{R}$, with $2n = 2$ or 4, be a Hamiltonian which is equal to a hyperbolic quadratic form Q at infinity such that Q has only real* *eigenvalues. Assume that φ_H is strongly nondegenerate and has at least two fixed points, and* $\mathrm{Fix}(\varphi_H)$ *is finite. Then φ_H has simple periodic orbits of arbitrarily large period.*

Note that strong nondegeneracy is a C^∞-generic condition in the class of Hamiltonians under consideration. Let us also point out that, in contrast with many closed manifolds (see, e.g., [Ginzburg and Gürel 2009b]), the existence of infinitely many periodic orbits is obviously not a C^∞- or even C^2-generic property of Hamiltonians in Theorem 1.4: one has to have an extra periodic orbit which serves as a seed eventually "spawning an infinitude of offspring".

In dimension two, the strong nondegeneracy requirement can be relaxed. It suffices to just assume that φ_H has at least two isolated homologically nontrivial fixed points; see Theorem 4.5. (Also note that in this case the eigenvalues of a hyperbolic quadratic form are automatically real.) However in dimension four, nondegeneracy enters the proof in a crucial way. Finally note that the two-dimensional case of Theorem 1.4 is intimately related to Franks's theorem; see Remarks 4.6 and 4.7.

Remark 1.5. A more general version of Theorem 1.4 for $2n = 2$ was proved in [Abbondandolo 2001, Theorem 5.1.9].

Organization of the paper. In Section 2, we set conventions and notation, and briefly recall some of the tools used in the paper and provide relevant references. We establish a version of the maximum principle and show that the Floer homology, as well as the relevant continuation maps, are defined for the class of Hamiltonians in question in Section 3. Finally in Section 4, we prove Theorems 1.1 and 1.4.

2. Conventions and notation

Throughout the paper, we will be working with the symplectic manifold $(\mathbb{R}^{2n}, \omega)$, where ω is the standard symplectic form. All Hamiltonians H considered here are assumed to be one-periodic in time, that is, $H : S^1 \times \mathbb{R}^{2n} \to \mathbb{R}$, and we set $H_t = H(t, \cdot)$ for $t \in S^1 = \mathbb{R}/\mathbb{Z}$. The Hamiltonian vector field X_H of H is defined by $i_{X_H}\omega = -dH$. The (time-dependent) flow of X_H is denoted by φ_H^t and its time-one map by φ_H. Such time-one maps are referred to as *Hamiltonian diffeomorphisms*. The action of a one-periodic Hamiltonian H on a loop $\gamma : S^1 \to \mathbb{R}^{2n}$ is defined by

$$\mathcal{A}_H(\gamma) = -\int_z \omega + \int_{S^1} H_t(\gamma(t))\, dt,$$

where $z : D^2 \to M$ is such that $z|_{S^1} = \gamma$. The least action principle asserts that the critical points of $\mathcal{A}_H$ on the space of all smooth maps $\gamma : S^1 \to \mathbb{R}^{2n}$ are exactly the one-periodic orbits of φ_H^t.

Let K and H be two one-periodic Hamiltonians. The *composition* $K \natural H$ is defined by the formula

$$(K \natural H)_t = K_t + H_t \circ (\varphi_K^t)^{-1}, \tag{2-1}$$

and the flow of $K \natural H$ is $\varphi_K^t \circ \varphi_H^t$. We set $H^{\natural k} = H \natural \cdots \natural H$ (k times). Abusing terminology, we will refer to $H^{\natural k}$ as the k-th iteration of H. Clearly $H^{\natural k} = kH$ when H is autonomous. (Note that the flow $\varphi_{H^{\natural k}}^t = (\varphi_H^t)^k$, $t \in [0, 1]$, is homotopic with fixed end-points to the flow φ_H^t, $t \in [0, k]$. Also, in general, $H^{\natural k}$ is not one-periodic, even when H is.) Furthermore, setting

$$\|F\|_B = \int_{S^1} \sup_B |F| \, dt \tag{2-2}$$

for a bounded set $B \subset \mathbb{R}^{2n}$, we have $\|H^{\natural k}\|_B = k\|H\|_B$ when H is autonomous. Note that $\|F\|_B$ is a variant of the Hofer norm. (When F is compactly supported on $\mathbb{R}^{2n}$, we will also use the notation $\|F\|_{\mathbb{R}^{2n}}$ with the obvious meaning.)

The k-th iteration of a one-periodic orbit γ of H will be denoted by γ^k. More specifically, $\gamma^k(t) = \varphi_{H^{\natural k}}^t(\gamma(0))$, where $t \in [0, 1]$. We can think of γ^k as the k-periodic orbit $\gamma(t)$, $t \in [0, k]$, of H. Hence there is an action-preserving one-to-one correspondence between one-periodic orbits of $H^{\natural k}$ and k-periodic orbits of H.

The *action spectrum* $\mathcal{S}(H)$ of H is the set of critical values of $\mathcal{A}_H$. This is a zero measure, closed (hence nowhere dense) set; see, for example, [Hofer and Zehnder 1994]. Clearly the action functional is homogeneous with respect to iteration:

$$\mathcal{A}_{H^{\natural k}}(\gamma^k) = k\mathcal{A}_H(\gamma).$$

A periodic orbit γ of H is said to be *nondegenerate* if the linearized return map $d\varphi_H : T_{\gamma(0)}M \to T_{\gamma(0)}M$ has no eigenvalues equal to one. A Hamiltonian is called nondegenerate if all its one-periodic orbits are nondegenerate and strongly nondegenerate if all k-periodic orbits (for all k) are nondegenerate.

Let γ be a nondegenerate periodic orbit. The *Conley–Zehnder index*, denoted by $\mu_{\mathrm{CZ}}(H, \gamma) \in \mathbb{Z}$, is defined, up to a sign, as in [Salamon 1999; Salamon and Zehnder 1992]. (When H is clear from the context we use the notation $\mu_{\mathrm{CZ}}(\gamma)$.) More specifically, in this paper, the Conley–Zehnder index is the negative of that in [Salamon 1999]. In other words, we normalize μ_{CZ} so that $\mu_{\mathrm{CZ}}(\gamma) = n$ when γ is a nondegenerate maximum of an autonomous Hamiltonian with small Hessian. Furthermore recall that the mean index $\Delta_H(\gamma)$ is defined regardless of whether γ is degenerate or not, and $\Delta_H(\gamma)$ depends continuously on H and γ in the obvious sense. When γ is nondegenerate, we have

$$0 \le |\Delta_H(\gamma) - \mu_{\mathrm{CZ}}(H, \gamma)| < n.$$

Furthermore the mean index is also homogeneous with respect to iteration:

$$\Delta_{H^{\natural k}}(\gamma^k) = k\Delta_H(\gamma).$$

3. Maximum principle and Floer homology

Our goal in this section is to show that the Floer homology is defined and has the standard properties for the class of Hamiltonians in question. In our setting, essentially the only issue to deal with is the compactness of moduli spaces of Floer trajectories, which we establish by proving a version of the maximum principle.

Floer homology. Let Q be a hyperbolic quadratic form on $\mathbb{R}^{2n}$, that is, Q is nondegenerate and has no eigenvalues on $i\mathbb{R}$. (Recall that throughout the paper by eigenvalues of Q we mean the eigenvalues of the linear Hamiltonian vector field X_Q.) Assume further that all eigenvalues of Q are real. (See Remark 3.3 for a variant of the maximum principle when Q has complex eigenvalues.) Denote by $\mathcal{H}_Q$ the set of one-periodic Hamiltonians $H : S^1 \times \mathbb{R}^{2n} \to \mathbb{R}$ which are compactly supported time-dependent perturbations of Q. Let $J = J_t$ be a time-dependent almost complex structure compatible with ω. We are interested in solutions $u : \mathbb{R} \times S^1 \to \mathbb{R}^{2n}$ of the Floer equation

$$\partial_s u + J(u)\partial_t u = -\nabla H_t(u), \tag{3-1}$$

where $u = u(s, t)$ with coordinates (s, t) on $\mathbb{R} \times S^1$ and the gradient is taken with respect to the one-periodic in time metric $\langle \cdot, \cdot \rangle = \omega(\cdot, J\cdot)$ on $\mathbb{R}^{2n}$.

In this setting we have:

Theorem 3.1. *Let Q be a hyperbolic quadratic form on $(\mathbb{R}^{2n}, \omega)$ with only real eigenvalues. Then there exists a linear complex structure J_Q compatible with ω such that whenever $J \equiv J_Q$ and $H \equiv Q$ outside an open ball B with respect to the metric $\langle \cdot, \cdot \rangle_Q := \omega(\cdot, J_Q \cdot)$, any solution of* (3-1) *for the pair (H, J) that is asymptotic to periodic orbits of H in B is necessarily contained in B.*

More generally, consider now solutions $u : \Omega \to \mathbb{R}^{2n}$ of (3-1), where $\Omega \subset \mathbb{R} \times S^1$ is an open connected subset. Theorem 3.1 is an immediate consequence of the following proposition:

Proposition 3.2 (maximum principle). *Let Q be a hyperbolic quadratic form on $(\mathbb{R}^{2n}, \omega)$ with only real eigenvalues. Then there exists a linear complex structure J_Q compatible with ω such that for any solution u (with domain $\Omega \subset \mathbb{R} \times S^1$) of the Floer equation* (3-1) *for (Q, J_Q), the function $\rho = \|u\|^2/2$, where the norm is induced by the metric $\langle \cdot, \cdot \rangle = \omega(\cdot, J_Q \cdot)$, cannot attain a maximum at an interior point of Ω unless ρ is constant.*

Proof of Proposition 3.2. Below we first introduce J_Q and then prove that ρ is subharmonic on Ω, that is, $\Delta\rho \geq 0$, where the Laplacian is taken with respect to metric $\omega(\cdot\,, J_Q\cdot\,)$.

Since we will be changing the basis and the inner product on the ambient space throughout the proof, it is more convenient to work with a hyperbolic quadratic form Q on a finite-dimensional symplectic vector space (V^{2n}, ω). Equip V with a symplectic basis

$$(\partial_p, \partial_q) = (\partial_{p_1}, \dots, \partial_{p_n}, \partial_{q_1}, \dots, \partial_{q_n})$$

such that in the corresponding coordinates (p, q) on $\mathbb{R}^{2n}$, the quadratic form Q is expressed as

$$Q(p, q) = \langle Ap, q\rangle. \tag{3-2}$$

Here A is a nondegenerate lower triangular $n \times n$ matrix, and $\langle\cdot\,,\cdot\rangle$ is the standard inner product on $\mathbb{R}^n$. Indeed, since Q is nondegenerate with only real eigenvalues, it can be expressed in some symplectic basis (∂_p, ∂_q) as the direct sum of the normal forms

$$\sigma \sum_{i=1}^{m} p_i q_i - \sum_{i=1}^{m-1} p_i q_{i+1},$$

where σ ranges over the positive eigenvalues of Q and m is the multiplicity of σ; see [Arnold 1978; Williamson 1936]. We emphasize that p's and q's are treated here as vectors in $\mathbb{R}^n$ using the bases (∂_p) and (∂_q), respectively. (It is clear from this formula that A is indeed lower triangular.) Note that with this choice all diagonal entries of A, that is, the eigenvalues of A, are positive.

Let $A = D + E$, where D is the diagonal part of A and E is the strictly lower triangular part. By rescaling the basis vectors (∂_p, ∂_q), while still keeping the basis symplectic and keeping (3-2), we can make E arbitrarily small. (We will specify shortly how small E has to be. Here we merely note that the rescaling does not affect D and that, in fact, E is required to be small compared to D.) We keep the notation (∂_p, ∂_q) for the new basis and (p, q) for the resulting linear coordinates.

The complex structure J_Q is defined by the requirement $J_Q\partial_p = -\partial_q$. This structure is compatible with ω, and we denote by $\langle\cdot\,,\cdot\rangle_Q$ the resulting inner product $\omega(\cdot\,, J_Q\cdot\,)$ on V, that is, $\langle\cdot\,,\cdot\rangle_Q := \omega(\cdot\,, J_Q\cdot\,)$. From now on we identify (V, ω) with the standard symplectic $\mathbb{R}^{2n}$ using the basis (∂_p, ∂_q). Under this identification, J_Q becomes the standard complex structure on $\mathbb{R}^{2n} = \mathbb{C}^n$, and $\langle\cdot\,,\cdot\rangle_Q$ turns into the standard inner product. Note also that the restriction of $\langle\cdot\,,\cdot\rangle_Q$ to the subspaces generated by ∂_p and ∂_q, respectively, is the inner product $\langle\cdot\,,\cdot\rangle$ on $\mathbb{R}^n$. Finally we emphasize that all these structures, except for ω, depend on the choice of the basis (∂_p, ∂_q) which is to be finalized below (after we state how small E needs to be).

In what follows, we calculate the Laplacian with respect to the metric $\langle \cdot, \cdot \rangle_Q$, where we set $u(s,t) = (p,q)$ and use the Floer equation (3-1):

$$
\begin{aligned}
\Delta\rho &= \rho_{ss} + \rho_{tt} \\
&= \|u_s\|^2 + \|u_t\|^2 - \langle u, \partial_s \nabla Q(u)\rangle_Q + \langle u, J_Q \partial_t \nabla Q(u)\rangle_Q \\
&= \|u_s\|^2 + \|u_t\|^2 + \langle A^2 p, p\rangle + \langle A^2 q, q\rangle \\
&\qquad + \langle p, (A - A^T) q_s\rangle - \langle q, (A - A^T) p_s\rangle \\
&= \|u_s\|^2 + \|u_t\|^2 + \|Dp\|^2 + \|Dq\|^2 + \langle E^2 p, p\rangle + \langle E^2 q, q\rangle \\
&\qquad + \langle (DE + ED)\, p, p\rangle + \langle (DE + ED)\, q, q\rangle \\
&\qquad + \langle p, (E - E^T) q_s\rangle - \langle q, (E - E^T) p_s\rangle.
\end{aligned}
\tag{3-3}
$$

Next we specify the requirements on E. To this end, let $\lambda := \min \lambda_i > 0$, where the λ_i are the eigenvalues of A (or D). Then we have

$$\|Dx\|^2 \geq \lambda^2 \|x\|^2 \quad \text{for any } x \in \mathbb{R}^n. \tag{3-4}$$

Now E is required to be so small that:

(i) $|\langle E^2 x, x\rangle| \leq \lambda^2 \|x\|^2/10$ for any $x \in \mathbb{R}^n$,

(ii) $|\langle DE\, x, x\rangle| \leq \lambda^2 \|x\|^2/20$ and $|\langle ED\, x, x\rangle| \leq \lambda^2 \|x\|^2/20$ for any $x \in \mathbb{R}^n$,

(iii) $|\langle x, (E - E^T) y\rangle| \leq \lambda \|x\|\, \|y\|/8$ for any x and $y \in \mathbb{R}^n$.

Using (3-4), and (i) and (ii) for $x = p$ and $x = q$, and (iii) for $(x,\ y) = (p,\ q_s)$ and $(x,\ y) = (q,\ p_s)$ in (3-3), it is straightforward to show that

$$
\begin{aligned}
\Delta\rho &\geq \frac{3\lambda^2}{10}\|u\|^2 + \|u_s\|^2 + \frac{\lambda^2}{2}\|u\|^2 - \frac{\lambda}{4}\|u\|\,\|u_s\| \\
&\geq \frac{3\lambda^2}{10}\|u\|^2 + \left(\|u_s\| - \frac{\lambda}{\sqrt{2}}\|u\|\right)^2 \geq \frac{3\lambda^2}{10}\|u\|^2 \geq 0.
\end{aligned}
\tag{3-5}
$$

□

Remark 3.3. It is not hard to see that Proposition 3.2 still holds when the quadratic form Q has complex eigenvalues σ, provided that $|\mathrm{Re}\,\sigma| > |\mathrm{Im}\,\sigma|$ or, equivalently, $\mathrm{Re}\,\sigma^2 > 0$ for all eigenvalues. However in general without this assumption (or when Q is elliptic but not positive definite), there seems to be no reason to expect the maximum principle to hold. There are also several other variants of the maximum principle which hold for solutions of the Floer equation for Q. For instance, it holds for the functions $\|p\|^2$ and $\|q\|^2$ separately.

As a consequence of Theorem 3.1, the total and filtered Floer homology groups of $H \in \mathcal{H}_Q$, denoted by $\mathrm{HF}(H)$ and $\mathrm{HF}^{(a,b)}(H)$, respectively, are defined and have properties similar to those for closed symplectically aspherical manifolds; see, for example, [Hofer and Zehnder 1994; McDuff and Salamon 2004]. (For the sake of simplicity all homology groups are taken over $\mathbb{Z}_2$.) Likewise the local Floer homology $\mathrm{HF}(H,\gamma)$ of H at an isolated periodic orbit γ is also defined and has the usual properties; see, for example, [Floer 1989a; 1989b; Ginzburg 2010; Ginzburg and Gürel 2010]. (Here J is an ω-compatible almost complex structure which is generic within the class of almost complex structures equal outside a compact set to some J_Q, as in Theorem 3.1. We will discuss the dependence of the Floer homology on J shortly.)

Our next goal is to define the continuation maps induced by homotopies of Hamiltonians in $\mathcal{H}_Q$. To this end, we say that a homotopy $F_s = Q + f_s$ in $\mathcal{H}_Q$ from H_0 to H_1 is compactly supported if $\bigcup_s \operatorname{supp} f_s$ is bounded. (Observe that this is not automatically the case.) Then we have a continuation map

$$\Psi : \mathrm{HF}^{(a,\,b)}(H_0) \to \mathrm{HF}^{(a,\,b)+C}(H_1) \tag{3-6}$$

for any H_0 and H_1 in $\mathcal{H}_Q$, induced by a homotopy F_s in $\mathcal{H}_Q$. Here $(a,\,b)+C$ stands for $(a+C,\,b+C)$, and

$$C \geq \int_{-\infty}^{\infty}\int_{S^1} \sup\nolimits_{\mathbb{R}^{2n}} \partial_s F_s\,dt\,ds = \int_{-\infty}^{\infty}\int_{S^1} \sup\nolimits_{\mathbb{R}^{2n}} \partial_s f_s\,dt\,ds, \tag{3-7}$$

with $\partial_s F_s \equiv 0$ when $|s|$ is large; see [Ginzburg 2007, Section 3.2.2]. Note that the suprema in (3-7) exist since the homotopy is compactly supported.

We now have $\mathrm{HF}(H) = \mathrm{HF}(Q) \cong \mathbb{Z}_2$, concentrated in degree $\mu_{\mathrm{CZ}}(Q,0)=0$. It is clear that the filtered Floer homology $\mathrm{HF}^{(a,\,b)}(H)$ is independent of the almost complex structure J as long as J_Q is fixed. However it is not obvious at all whether this homology is independent of the choice of J_Q. In what follows, we will always have J_Q fixed and suppress this hypothetical dependence in the notation.

Continuation maps beyond $\mathcal{H}_Q$. The class $\mathcal{H}_Q$ is not closed under iteration. For instance, $H^{\natural 2} \in \mathcal{H}_{2Q}$ when $H \in \mathcal{H}_Q$. To incorporate iteration into the picture, we consider a broader class $\widehat{\mathcal{H}}_Q$ which is the union of the classes $\mathcal{H}_{kQ}$ for all real $k>0$. Clearly this class is now closed under iteration. Moreover one can see from the proof of Proposition 3.2 that there exists a common almost complex structure, J_Q, for which Theorem 3.1 holds for all Hamiltonians in $\widehat{\mathcal{H}}_Q$ or, to be more precise, any J_Q can also be taken as J_{kQ}. (The reason is that J_Q is determined by the requirement that the off-diagonal part E of A is small *compared to D*, rather than just small. Thus if conditions (i), (ii) and (iii) are satisfied for Q, they are also automatically satisfied in the same basis for kQ for any $k>0$.) From now on we fix J_Q.

As above, we say that a homotopy $F_s = k(s)Q + f_s$ in $\widehat{\mathcal{H}}_Q$ is compactly supported if $\bigcup_s \operatorname{supp} f_s$ is bounded and call the closure of this union the support of the homotopy. A homotopy is called *slow* if it is compactly supported and, say,

$$\frac{|k'(s)|}{(k(s))^2} \le \frac{3\lambda^2}{20} \inf_{x \in \mathbb{R}^{2n}\setminus\{0\}} \frac{\|x\|^2}{|Q(x)|}, \tag{3-8}$$

where $\lambda = \min \lambda_i$ is as in Section 3; cf. [Cieliebak et al. 1995]. Clearly the right-hand side in (3-8) is positive. (In fact, the infimum in (3-8) is equal to $1/\lambda_{\max}$, where $\lambda_{\max}$ is the largest of the absolute values of the eigenvalues of Q with respect to $\|x\|^2$. This follows from the Courant–Fischer minimax theorem; see, for example, [Demmel 1997, Chapter 5].) Recall also that a homotopy F_s from H_0 to H_1 is called linear if $F_s = (1-g(s))H_0 + g(s)H_1$, where $g : \mathbb{R} \to \mathbb{R}$ is an increasing smooth function equal to zero for $s \ll 0$ and one for $s \gg 1$.

Theorem 3.4. *Let F_s be a slow homotopy in $\widehat{\mathcal{H}}_Q$ from H_0 to H_1, supported in a ball B with respect to the metric $\langle\cdot,\cdot\rangle_Q = \omega(\cdot, J_Q\cdot)$. The continuation map Ψ as in* (3-6) *is defined*, *where C satisfies*

$$C \ge \int_{-\infty}^{\infty}\int_{S^1} \sup\nolimits_B \partial_s F_s \, dt\, ds. \tag{3-9}$$

This map is independent of the slow homotopy. Furthermore, *for a linear slow homotopy*, *we can take $C = \|H_1 - H_0\|_B := \int_{S^1} \sup_B |H_1 - H_0|\, dt$.*

It is not hard to see that any compactly supported homotopy in $\widehat{\mathcal{H}}_Q$ can be reparametrized to make it slow without changing the right-hand side in (3-9). Note also that although the notion of a slow homotopy is independent of the size of the support, the lower bound in (3-9) does depend in general on the ball B containing the support and increases with the size of B. However one can show that the continuation map Ψ is independent of the ball B in the following sense: whenever for a fixed homotopy and two different balls C satisfies (3-9) for both of the balls, the resulting continuation map is independent of the ball. (In what follows, we will not use this fact.)

The continuation maps Ψ have properties similar to their counterparts in the ordinary Floer homology. For instance, the continuation map induced by a concatenation of homotopies is equal to the composition of the continuation maps, and continuation maps commute with the maps in the long exact sequence in filtered Floer homology. (See [Ginzburg 2007] for a detailed account on the so-called C-bounded homotopies in filtered Floer homology.) Note however that here, as in (3-6), the almost complex structure J_s is independent of s outside B.

Remark 3.5. We emphasize that the continuation map Ψ is not necessarily defined when the homotopy is not slow.

Proof of Theorem 3.4. To prove that Ψ is well-defined, it suffices to show that the maximum principle, Proposition 3.2, extends to solutions of the Floer equation for slow homotopies $Q_s = k(s)Q$ connecting $Q_0 = k_0 Q$ and $Q_1 = k_1 Q$, where $k_0 = k(0)$ and $k_1 = k(1)$. To this end, note that $Q_s = k(s)\langle Ap, q\rangle$ and recall that, as was noted above, we can take J_{kQ} to be J_Q. Calculating the Laplacian with respect to the metric $\langle\cdot,\cdot\rangle_Q = \omega(\cdot, J_Q\,\cdot)$ in this setting, we obtain

$$\begin{aligned}\Delta\rho &= \|u_s\|^2 + \|u_t\|^2 + k(s)\big(\langle p, (A-A^T)q_s\rangle - \langle q, (A-A^T)p_s\rangle\big)\\ &\qquad + (k(s))^2(\langle A^2 p, p\rangle + \langle A^2 q, q\rangle) - 2k'(s)\langle Ap, q\rangle\\ &= \|u_s\|^2 + \|u_t\|^2 + (k(s))^2(\|Dp\|^2 + \|Dq\|^2 + \langle E^2 p, p\rangle + \langle E^2 q, q\rangle)\\ &\qquad + (k(s))^2\big(\langle (DE+ED)\,p, p\rangle + \langle (DE+ED)\,q, q\rangle\big)\\ &\qquad + k(s)\big(\langle p, (E-E^T)q_s\rangle - \langle q, (E-E^T)p_s\rangle\big) - 2k'(s)\langle Ap, q\rangle\\ &\geq \frac{3\lambda^2}{10}(k(s))^2\|u\|^2 + \left(\|u_s\| - \frac{\lambda|k(s)|}{\sqrt{2}}\|u\|\right)^2 - 2|k'(s)|\cdot|Q(u)|.\end{aligned}$$

Hence $\Delta\rho \geq 0$ by (3-8).

That we can take C satisfying (3-9) for a general slow homotopy and that $C = \|H_1 - H_0\|_B$ satisfies (3-9) for a linear slow homotopy is established by a standard calculation (see, e.g., [Ginzburg 2007; Schwarz 2000]) combined with the observation that homotopy trajectories (i.e., solutions of (3-1) for the pair (F_s, J_s)) are confined to B due to the maximum principle. □

Remark 3.6. The maximum principle is also known to hold for positive definite quadratic Hamiltonians; see [McDuff 1991; Viterbo 1999] and also [Seidel 2008]. This fact underlies the definition of symplectic homology, and in fact, it was the motivation of our approach in this paper. However it is worth pointing out that in this case the continuation map between a Hamiltonian equal to kQ at infinity and the one equal to $(k+1)Q$ at infinity is defined only in one direction and this map, depending on Q, may be zero. This is the main reason why our approach to the proof of Theorem 1.1 does not carry over to positive definite quadratic Hamiltonians.

4. Proofs and generalizations

Proof of Theorem 1.1. As has been mentioned in the introduction, we establish a more general result. To state it, recall again that an isolated periodic orbit x is said to be homologically nontrivial if the local Floer homology of H at x is nonzero. For instance, a nondegenerate fixed point is homologically nontrivial. More generally, an isolated fixed point with nonvanishing topological index is homologically nontrivial, for this index is equal, up to a sign, to the Euler characteristic of the local Floer homology. The notion of homological nontriviality seems to be particularly well suited for use in the context of HZ and Conley conjectures; see, for example, Remark 1.3. Theorem 1.1 is an immediate consequence of the following result.

Theorem 4.1. *Let $H : S^1 \times \mathbb{R}^{2n} \to \mathbb{R}$ be a Hamiltonian which is equal to a hyperbolic quadratic form Q at infinity* (*i.e.*, *outside a compact set*) *such that Q has only real eigenvalues. Assume that φ_H has an isolated homologically nontrivial fixed point x with nonzero mean index and* $\mathrm{Fix}(\varphi_H)$ *is finite. Then φ_H has simple periodic orbits of arbitrarily large period.*

Proof. In what follows, for the sake of brevity, we suppress the t-dependence when taking a supremum or specifying the support of a function. For instance, when we say that a function is supported in $Y \subset \mathbb{R}^{2n}$, we mean that the support is in $S^1 \times Y$. Likewise two functions being equal on Y means that they are equal on $S^1 \times Y$, etc. Finally the supremum, without a set specified, will stand for the supremum over $\mathbb{R}^{2n}$.

Let $H = Q + f$ as in the statement of the theorem. Pick a polyball $P = B^n \times B^n$ containing $\operatorname{supp} f$ and a ball $V \supset P$. Throughout the proof, as in Section 3, we assume that the off-diagonal part E of A is small enough when compared to the diagonal part D. In particular, every integral curve of the flow of Q intersects P along a connected set. Before we actually turn to the proof of the theorem, we need to first modify H, without essentially changing its dynamics, to control the energy shift resulting from the homotopy between different iterations of H.

Lemma 4.2. *There exist constants $C_1 > 0$ and $C_2 > 0$, depending only on the quadratic form Q and the ball V, such that for every $\epsilon \in (0, 1]$, there exists an autonomous Hamiltonian $\widetilde{Q}$ with the following properties*:

(i) $\widetilde{Q} = Q$ *on* V.

(ii) $\widetilde{Q} = \epsilon Q$ *outside a ball* $V_\epsilon \supset V$ *of radius* $R = C_1/\sqrt{\epsilon}$.

(iii) $\sup_{V_\epsilon} |\widetilde{Q}| = C_2$.

(iv) *The Hamiltonian flow of $\widetilde{Q}$ has no periodic orbits other than the origin*, *and every integral curve of its flow intersects P along a connected set.*

The essential point here is that the constants C_1 and C_2 are independent of ϵ while $R = C_1/\sqrt{\epsilon}$ (but not, say, of order $1/\epsilon$). We will prove this lemma by giving an explicit construction of $\widetilde{Q}$ after the proof of Theorem 4.1. One can think of $\widetilde{Q}$ as a family of Hamiltonians smoothly parametrized by ϵ with $\widetilde{Q} = Q$ for $\epsilon = 1$.

Consider now the Hamiltonian

$$\widetilde{H} = \widetilde{Q} + f = \epsilon Q + (\widetilde{Q} - \epsilon Q) + f = \epsilon Q + h,$$

where $h = (\widetilde{Q} - \epsilon Q) + f$ is supported in V_ϵ. Observe that $\widetilde{H} \in \widehat{\mathcal{H}}_Q$. Furthermore $\widetilde{H} = H$ in V, the ball where the Hamiltonians have nontrivial dynamics. Moreover for every period, $\widetilde{H}$ and H have exactly the same periodic orbits by Lemma 4.2(iv), and the orbits have the same actions and indices. In fact, one might expect these Hamiltonians to have exactly the same filtered Floer homology with isomorphism induced by a slow linear homotopy. However we have not been able to prove this fact.

The next lemma concerns the iterations $\widetilde{H}^{\natural k}$ and an estimate, independent of k, of the difference $\widetilde{H}^{\natural(k+\ell)} - \widetilde{H}^{\natural k}$, which will be essential for the proof of Theorem 4.1.

Lemma 4.3. *The Hamiltonian* $\widetilde{H}^{\natural k}$ *satisfies the following conditions*:

(i) $\widetilde{H}^{\natural k} \in \widehat{\mathcal{H}}_Q$ *and is equal to* $k\epsilon Q$ *outside the ball* B_k *of radius* $\|\varphi_Q^{\epsilon(k-1)}\| R$ *centered at the origin, where* $\varphi_Q^{\epsilon(k-1)}$ *is viewed as a linear operator.*

(ii) *Assume that* k, ℓ *and* ϵ *are such that* $\|\varphi_Q^{\epsilon(k+\ell-1)}\| \leq 2$. *Then*

$$\|\widetilde{H}^{\natural(k+\ell)} - \widetilde{H}^{\natural k}\|_{B_{k+\ell}} \leq C_3 \ell, \tag{4-1}$$

where C_3 *is independent of* k, ℓ *and* ϵ, *and the norm is as defined in* (2-2).

Proof of Lemma 4.3. Denote by B_1 the ball V_ϵ from Lemma 4.2, that is, B_1 is the ball of radius $R = C_1/\sqrt{\epsilon}$ centered at the origin. Consider the nested sets

$$Y_k = \bigcup_{t\in[0,\,1]} \varphi_{\epsilon Q}^{(k-1)t}(B_1) \quad \text{for } k \in \mathbb{N}.$$

Let $B_k = B(R_k)$ be the ball of radius $R_k = \|\varphi_{\epsilon Q}^{(k-1)}\| R$. Clearly $B_k \supset Y_k$.

Recall that $\widetilde{H} = \epsilon Q + h$, where $h = (\widetilde{Q} - \epsilon Q) + f$ is supported in B_1. Observe that $\widetilde{H}^{\natural k}$ can be expressed as

$$\widetilde{H}^{\natural k} = k\epsilon Q + \sum_{j=0}^{k-1} h \circ (\varphi_{\widetilde{H}}^t)^{-j} + \epsilon \sum_{j=0}^{k-1} \bigl(Q \circ (\varphi_{\widetilde{H}}^t)^{-j} - Q\bigr) = k\epsilon Q + h_k.$$

We now show that $\operatorname{supp} h_k \subset Y_k$, which settles (i). Since $\operatorname{supp} h \subset B_1$, a point x can be in $\operatorname{supp}(h \circ (\varphi_{\widetilde{H}}^t)^{-j})$ only if $(\varphi_{\epsilon Q}^\tau)^{-j}(x) \in B_1$ for some $\tau \in [0,\,t]$. This implies that

$$x \in ((\varphi_{\epsilon Q}^\tau)^{-j})^{-1}(B_1) = \varphi_{\epsilon Q}^{j\tau}(B_1) \subset \bigcup_{t\in[0,\,1]} \varphi_{\epsilon Q}^{jt}(B_1) = Y_{j+1}.$$

Hence the first term in h_k is supported in Y_k. Dealing with the second term in h_k, we first note that

$$\epsilon\bigl(Q \circ (\varphi_{\widetilde{H}}^t)^{-j} - Q\bigr) = \epsilon\bigl(Q \circ (\varphi_{\widetilde{H}}^t)^{-j} - Q \circ (\varphi_{\epsilon Q}^t)^{-j}\bigr)$$

since Q is autonomous. Now it is clear that $(\varphi_{\widetilde{H}}^t)^{-j}(x) \neq (\varphi_{\epsilon Q}^t)^{-j}(x)$ only when the integral curve of ϵQ through x for $[-jt,\,0]$ enters B_1, that is, $(\varphi_{\epsilon Q}^\tau)^{-j}(x) \in B_1$ for some $\tau \in [0,\,t]$. Hence, similarly to the first term, the second term in h_k is also supported in Y_k, and we have $\operatorname{supp} h_k \subset Y_k$.

To establish (ii), denote by $B(1)$ the unit ball and observe that

$$\begin{aligned}\sup_{B_{k+\ell}} \left|h\circ(\varphi^t_{\widetilde{H}})^{-j}\right| &= \sup_{B_1} |\widetilde{Q}-\epsilon Q+f| \\ &\le \sup_{B_1} |\widetilde{Q}|+\epsilon \sup_{B_1} |Q|+\sup|f| \\ &\le C_2+\epsilon\frac{C_1^2}{\epsilon}\sup_{B(1)}|Q|+\sup|f| \\ &= C_2+C_1^2\sup_{B(1)}|Q|+\sup|f|\end{aligned}$$

for any $j=k,\dots,k+\ell-1$. Furthermore, by the energy conservation law, we have

$$\begin{aligned}\sup_{B_{k+\ell}}\left|Q\circ(\varphi^t_{\widetilde{H}})^{-j}-Q\right| &= \sup_{B_{k+\ell}}\left|Q\circ(\varphi^t_{\widetilde{H}})^{-j}-Q\circ(\varphi^t_{\epsilon Q})^{-j}\right| \\ &= \sup_{B_1}\left|Q\circ(\varphi^t_{\widetilde{H}})^{-j}-Q\circ(\varphi^t_{\epsilon Q})^{-j}\right| \\ &\le 2\sup_{B_1}|Q|\le 2\frac{C_1^2}{\epsilon}\sup_{B(1)}|Q|.\end{aligned}$$

Now recall that k, ℓ and ϵ are such that $\|\varphi_Q^{\epsilon(k+\ell-1)}\|\le 2$. Thus $B_{k+\ell}\subset 2B_1$. Setting $M=\sup_{B(1)}|Q|$ and using the above estimates, we have

$$\begin{aligned}&\sup_{B_{k+\ell}}|\widetilde{H}^{\natural(k+\ell)}-\widetilde{H}^{\natural k}| \\ &\quad= \sup_{B_{k+\ell}}\left|\ell\epsilon Q+\sum_{j=k}^{k+\ell-1}h\circ(\varphi^t_{\widetilde{H}})^{-j}+\epsilon\sum_{j=k}^{k+\ell-1}\big(Q\circ(\varphi^t_{\widetilde{H}})^{-j}-Q\big)\right| \\ &\quad\le \ell\epsilon\sup_{2B_1}|Q| \\ &\qquad\quad+\sum_{j=k}^{k+\ell-1}\sup_{B_{k+\ell}}|h\circ(\varphi^t_{\widetilde{H}})^{-j}|+\epsilon\sum_{j=k}^{k+\ell-1}\sup_{B_{k+\ell}}\left|\big(Q\circ(\varphi^t_{\widetilde{H}})^{-j}-Q\big)\right| \\ &\quad\le 4\ell\epsilon\frac{C_1^2}{\epsilon}M+\ell(C_2+C_1^2M+\sup|f|)+2\ell\epsilon\frac{C_1^2}{\epsilon}M\le(7C_1^2M+C_2+\sup|f|)\ell.\end{aligned}$$

Setting $C_3:=7C_1^2M+C_2+\sup|f|$, we then have

$$\|\widetilde{H}^{\natural(k+\ell)}-\widetilde{H}^{\natural k}\|_{B_{k+\ell}}=\int_{S^1}\sup_{B_{k+\ell}}|\widetilde{H}^{\natural(k+\ell)}-\widetilde{H}^{\natural k}|\,dt\le C_3\ell, \tag{4-2}$$

with C_3 independent of k, ℓ and ϵ, as required. □

From now on we will work with the Hamiltonians $\widetilde{H}$, and at this stage we prefer not to specify the parameter ϵ yet. These Hamiltonians have the same periodic orbits with the same actions and indices, and up to the point when the homotopy between the iterated Hamiltonians is considered, the argument applies to any of the Hamiltonians $\widetilde{H}$.

It is worth mentioning again that the Hamiltonians $\widetilde{H}^{\natural k}$ are not one-periodic in time even though $\widetilde{H}$ is. This issue however is quite standard and can be dealt with in a straightforward way. Namely consider a Hamiltonian $G = K + g$, where $g = g_t$ is time-dependent for $t \in [0, 1]$ and K is any autonomous Hamiltonian. The Hamiltonian diffeomorphism φ_G can be generated by a one-periodic Hamiltonian

$$\bar{G} = K + \lambda'(t) g_{\lambda(t)} \circ \varphi_K^{\lambda(t)-t},$$

where $\lambda : [0, 1] \to [0, 1]$ is an increasing function equal to zero for $t \approx 0$ and one for $t \approx 1$. We apply this procedure to $\widetilde{H}^{\natural k}$ with $K = k\epsilon Q$ and $g = h_k$. The actions, the Conley–Zehnder indices and the mean indices of the periodic orbits do not change. The change of the set B_k can be made arbitrarily small, of the order $\|1 - \lambda'(t)\|_{L^1}$. As a consequence, the upper bound (4-1) can also be adjusted by an arbitrarily small amount independent of k. In what follows, we will treat the Hamiltonians $\widetilde{H}^{\natural k}$ as one-periodic in time, allowing for these straightforward modifications.

Now we are in a position to proceed with the proof of Theorem 4.1. It suffices to show that there exist arbitrarily large primes which occur as periods of simple periodic orbits. Arguing by contradiction, assume that only finitely many prime numbers are attained as the periods. From now on, we always denote by p or p_i a prime number greater than the largest period. Let $\widetilde{H} = \widetilde{Q} + f$, where $\widetilde{Q}$ is any Hamiltonian from Lemma 4.2. Then for any such prime p, all p-periodic orbits of $\varphi_{\widetilde{H}}$ are iterations of fixed points of φ_H, and hence $\mathcal{S}(\widetilde{H}^{\natural p}) = p\,\mathcal{S}(H)$. Recall in this connection that φ_H is assumed to have finitely many fixed points. Next let us note that all sufficiently large prime numbers are admissible in the sense of [Ginzburg and Gürel 2010]. Thus under such iterations of $\widetilde{H}$, the orbit x stays isolated, and

$$\mathrm{HF}(\widetilde{H}^{\natural p}, x^p) = \mathrm{HF}(H^{\natural p}, x) = \mathrm{HF}(H, x)$$

up to, in the second equality, a shift of degree determined by the order of iteration p; see [Ginzburg and Gürel 2010, Theorem 1.1]. In particular, in our case, $\mathrm{HF}(\widetilde{H}^{\natural p}, x^p) \neq 0$ since $\mathrm{HF}(H, x) \neq 0$.

As has been mentioned above, $\Delta_{\widetilde{H}}(x) = \Delta_H(x)$, and let us assume that $\Delta_H(x) > 0$, for the argument is similar if $\Delta_H(x) < 0$. Moreover let us assume for the sake of simplicity that $\mathcal{A}_H(x) = 0$ and hence $\mathcal{A}_{\widetilde{H}}(x) = 0$. (The general case can be dealt with in a similar fashion and requires only notational modifications.) Consequently $\mathcal{A}_{\widetilde{H}^{\natural p}}(x^p) = 0$ for all iterations p. Let $a > 0$ be outside $\mathcal{S}(\widetilde{H}) = \mathcal{S}(H)$ such that 0 is the only point in $(-a, a) \cap \mathcal{S}(H)$ and therefore in $(-ap, ap) \cap \mathcal{S}(\widetilde{H}^{\natural p})$. Then we have

$$\mathrm{HF}_*^{(-ap, ap)}(\widetilde{H}^{\natural p}) = \mathrm{HF}_*(H^{\natural p}, x^p) \oplus \cdots, \tag{4-3}$$

where the dots represent the local Floer homology contributions from the fixed points with zero action other than x. Furthermore we henceforth focus on degrees $*$

such that $|*| > n$. This guarantees that the fixed points with zero mean index do not contribute to $\mathrm{HF}_*^{(-ap,ap)}(\widetilde{H}^{\natural p})$, for their local Floer homology groups are supported in $[-n, n]$, where the support is by definition the set of degrees for which the local Floer homology groups are nonzero. Thus all terms on the right-hand side of (4-3) come from fixed points with nonzero mean index. Moreover we can further restrict $*$ so that only the fixed points γ having the same mean index as x contribute to the right-hand side of (4-3). This is possible since the supports of local Floer homology groups coming from fixed points with other nonzero mean indices are separated from $\operatorname{supp}\mathrm{HF}_*(H^{\natural p}, x^p) \subset [p\Delta_H(x)-n,\ p\Delta_H(x)+n]$ whenever p is sufficiently large.

From now on, we work with primes $p > 2$ which are as large as is needed above. Let us order these prime numbers as $p_1 < p_2 < \cdots$. In what follows, p_i always denotes a prime from this sequence.

Next recall that $\Delta_H(x) > 0$ and let $m \in \mathbb{N}$ be such that $m > n/\Delta_H(x)$. Then using the fact that $p_{i+m} - p_i \geq 2m$, we see that the supports of $\mathrm{HF}(H^{\natural p_i}, \gamma^{p_i})$ and $\mathrm{HF}(H^{\natural p_{i+m}}, \gamma^{p_{i+m}})$ are disjoint for all i and for all fixed points γ of φ_H with $\mathcal{A}_H(\gamma) = 0$ and $\Delta_H(\gamma) = \Delta_H(x)$. This is because

$$[p_i\Delta_H(x)-n,\ p_i\Delta_H(x)+n] \cap [p_{i+m}\Delta_H(x)-n,\ p_{i+m}\Delta_H(x)+n] = \varnothing,$$

where the first interval contains $\operatorname{supp}\mathrm{HF}(H^{\natural p_i}, \gamma^{p_i})$ and the second one contains $\operatorname{supp}\mathrm{HF}(H^{\natural p_{i+m}}, \gamma^{p_{i+m}})$. Moreover for any p_i, there exists an integer s_i such that $\mathrm{HF}_{s_i}(H^{\natural p_i}, x^{p_i}) \neq 0$, as is mentioned earlier and proved in [Ginzburg and Gürel 2010]. Thus we see that

$$\mathrm{HF}_{s_i}(H^{\natural p_i}, x^{p_i}) \neq 0 \quad \text{and} \quad \mathrm{HF}_{s_i}(H^{\natural p_{i+m}}, \gamma^{p_{i+m}}) = 0 \tag{4-4}$$

for all fixed points γ as above since s_i is outside $\operatorname{supp}\mathrm{HF}(H^{\natural p_{i+m}}, \gamma^{p_{i+m}})$ for all such γ.

Choose p_i so large that $p_i a > 6C_3(p_{i+m} - p_i)$, where C_3 is introduced in Lemma 4.3. That one can do so is guaranteed by the equality $p_{i+1} - p_i = o(p_i)$; see [Baker et al. 2001]. (Obviously one can write $p_{i+m} - p_i$ as a telescoping sum of the differences of two consecutive primes, and hence by a simple inductive argument, $p_{i+m} - p_i = o(p_i)$.) Now pick $\alpha > 0$, depending on m and i, such that

$$-p_i a < -\alpha < -\alpha + 2C_3(p_{i+m} - p_i) < 0 < \alpha < \alpha + 2C_3(p_{i+m} - p_i) < p_i a.$$

For instance, α satisfying $p_i a - 4C_3(p_{i+m} - p_i) < \alpha < p_i a - 2C_3(p_{i+m} - p_i)$ would work. As a consequence, we also have

$$-p_{i+m} a < -\alpha + C_3(p_{i+m} - p_i) < 0 < \alpha + C_3(p_{i+m} - p_i) < p_{i+m} a.$$

Finally let us specify $\widetilde{Q}$, and in turn $\widetilde{H}$. To this end, we choose $\epsilon > 0$ so small that $\|\varphi_Q^{\epsilon(p_{i+m}-1)}\| \leq 2$ and hence (4-1) is satisfied with $k = p_i$ and $k + l = p_{i+m}$

in the second assertion of Lemma 4.3. Set $\delta := C_3(p_{i+m} - p_i)$. Then for a linear homotopy, which we may assume to be slow in the sense of Section 3, from $\widetilde{H}^{\natural p_i}$ to $\widetilde{H}^{\natural p_{i+m}}$, we have the induced map

$$\mathrm{HF}^{(-\alpha,\,\alpha)}(\widetilde{H}^{\natural p_i}) \to \mathrm{HF}^{(-\alpha,\,\alpha)+\delta}(\widetilde{H}^{\natural p_{i+m}}).$$

Here the fact that δ is the correct action shift follows from Theorem 3.4 and Lemma 4.3. Likewise the linear-homotopy map from $\widetilde{H}^{\natural p_{i+m}}$ to $\widetilde{H}^{\natural p_i}$ results in another action shift in δ. Consider now the following commutative diagram:

$$\begin{tikzcd}
 & \mathrm{HF}_{s_i}^{(-\alpha,\,\alpha)+\delta}(\widetilde{H}^{\natural p_{i+m}}) = 0 \arrow[dr] & \\
0 \neq \mathrm{HF}_{s_i}^{(-\alpha,\,\alpha)}(\widetilde{H}^{\natural p_i}) \arrow[ur] \arrow[rr, "\cong"] & & \mathrm{HF}_{s_i}^{(-\alpha,\,\alpha)+2\delta}(\widetilde{H}^{\natural p_i})
\end{tikzcd}$$

Here the top group is zero due to our choice of the degree s_i. On the other hand,

$$\mathrm{HF}_{s_i}^{(-\alpha,\,\alpha)}(\widetilde{H}^{\natural p_i}) = \mathrm{HF}_{s_i}(H^{\natural p_i}, x^{p_i}) \oplus \cdots \neq 0,$$

and the horizontal arrow is induced by the natural quotient-inclusion map; see, for example, [Ginzburg 2007]. This is indeed an isomorphism by the stability of filtered Floer homology (see, e.g., [Ginzburg and Gürel 2010]) because 0 is the only action value in the intervals $(-\alpha,\, \alpha)$ and $(-\alpha,\, \alpha) + 2\delta$. To summarize, a nonzero isomorphism factors through a zero group in the diagram. This contradiction completes the proof of Theorem 4.1, modulo a proof of Lemma 4.2, which is given below. $\square$

Remark 4.4. Notice that we have actually established the existence of a simple periodic orbit of either $H^{\natural p_i}$ or $H^{\natural p_{i+m}}$. In particular, starting with a sufficiently large prime number, among every m consecutive primes, there exists at least one prime which is the period of a simple periodic orbit of φ_H.

Furthermore for an infinite sequence of simple p_l-periodic orbits x_l of φ_H found this way, where $p_l \to \infty$, we have $\Delta_{H^{\natural p_l}}(x_l)/p_l \to \Delta_H(x)$. Hence in some sense, the mean index $\Delta_H(x)$ is an accumulation point in the union of normalized index spectra for H and its all iterations. (Of course, $\Delta_H(x)$ could possibly be isolated, but then $\Delta_{H^{\natural p_l}}(x_l)/p_l = \Delta_H(x)$.) A similar fact also holds for the action.

Finally note that the condition that the eigenvalues σ of Q are real can be relaxed and replaced by the requirement that $|\mathrm{Re}\,\sigma| > |\mathrm{Im}\,\sigma|$; cf. Remark 3.3.

Proof of Lemma 4.2. We construct the function $\widetilde{Q}$ in three steps and then show that $\widetilde{Q}$ has the required properties.

Step 1. Set $c = \sup_V |Q|$. Let $\eta : \mathbb{R} \to \mathbb{R}$ be an *odd* smooth function that equals the identity in $[-c, c]$ and satisfies $\eta' \geq \epsilon/2$ everywhere and

$$\eta(x) = \epsilon x \quad \text{when } |x| \geq c' = 2c/\epsilon.$$

It is easy to see that such a function exists. Note that to have a monotone function η such that $\eta(x) = x$ when $|x| \le c$ and $\eta(x) = \epsilon x$ when $|x| \ge c'$, we must have $c' > c/\epsilon$. This is the main reason why the radius R in the statement of the lemma must be of order $1/\sqrt{\epsilon}$. To start off the construction of $\widetilde{Q}$, we replace Q by $\eta \circ Q$.

Step 2. In the second step, we appropriately cut off $\eta \circ Q$ and define a new Hamiltonian $\widehat{Q}$ which is a linear transition from $\eta \circ Q$ to ϵQ in the q-direction. Namely let $r > 0$ be the radius of the ball V and set $a_0 = r/\sqrt{\epsilon}$ and $a_1 = 2a_0$. (So $r < a_0 < a_1$.) The modification of $\eta \circ Q$ takes place on the domain $a_0 \le \|q\| \le a_1$. To this end, choose a smooth monotone increasing function $\phi : [0, \infty) \to \mathbb{R}$ such that $\phi(x) = 0$ when $x \le a_0$, $\phi(x) = 1$ when $x \ge a_1$, and $|\phi'| \le 2/|a_1 - a_0| = 2/a_0$. We then set

$$\widehat{Q} = \phi(\|q\|)\epsilon Q + \big(1 - \phi(\|q\|)\big)\eta \circ Q.$$

Step 3. In the third step, we suitably cut off $\widehat{Q}$ and finally define the desired Hamiltonian $\widetilde{Q}$. To this end, let $b_0 = \max\{r, 32c/\lambda r\sqrt{\epsilon}\}$ and $b_1 = 2b_0$. Here, as in Section 3, $\lambda = \min \lambda_i > 0$, where λ_i's are the eigenvalues of A. (The reason for this choice of b_0 will be clear at the end of the proof.) Choose a smooth monotone increasing function $\psi : [0, \infty) \to \mathbb{R}$ such that $\psi(x) = 0$ when $x \le b_0$ and $\psi(x) = 1$ when $x \ge b_1$. Define

$$\widetilde{Q} = \psi(\|p\|)\epsilon Q + \big(1 - \psi(\|p\|)\big)\widehat{Q}.$$

Checking conditions (i)–(iv). Since $\eta \circ Q = Q$ on V by the definition of c, and $a_0 \ge r$ and $b_0 \ge r$, we clearly have $\widetilde{Q} = Q$ on the ball V. Furthermore $\widetilde{Q} = \epsilon Q$ outside the ball of radius $R = \sqrt{a_1^2 + b_1^2}$. Let V_ϵ be this ball. It is clear from our choice of a_1 and b_1 that R has the form $C_1/\sqrt{\epsilon}$, where C_1 is independent of ϵ. This proves (i) and (ii).

To establish (iii), observe first that

$$\begin{aligned} \sup|\eta(x) - \epsilon x| &= \sup\nolimits_{[0,c']}|\eta(x) - \epsilon x| \le \eta(c') + \epsilon c' \\ &= 2\epsilon c' = 4\epsilon c/\epsilon = 4c. \end{aligned} \tag{4-5}$$

Thus

$$\begin{aligned} \sup\nolimits_{V_\epsilon}|\widetilde{Q}| &\le \sup\nolimits_{V_\epsilon}|\epsilon Q| + \sup\nolimits_{V_\epsilon}|\widehat{Q}| \le 2\sup\nolimits_{V_\epsilon}|\epsilon Q| + \sup\nolimits_{V_\epsilon}|\eta \circ Q| \\ &\le 3\sup\nolimits_{V_\epsilon}|\epsilon Q| + \sup\nolimits_{V_\epsilon}|\eta \circ Q - \epsilon Q| \\ &\le 3\epsilon\frac{C_1^2}{\epsilon}\sup\nolimits_{B(1)}|Q| + 4c = 3C_1^2\sup\nolimits_{B(1)}|Q| + 4c =: C_2, \end{aligned}$$

with C_2 independent of ϵ. This is where replacing Q by $\eta \circ Q$ in Step 1 is essential.

To verify condition (iv), note that without loss of generality we may assume that the off-diagonal part of A is so small that

$$\pounds_{X_Q}\|p\|^2 \le -\lambda\|p\|^2 \quad \text{and} \quad \pounds_{X_Q}\|q\|^2 \le -\lambda\|q\|^2.$$

(Here we dropped the factor of 2 on the right-hand side of the inequalities to account for the off-diagonal terms.) In particular, every integral curve of Q enters the polyball P through the "side" part, $\|p\| = \text{const}$, of the boundary ∂P and leaves it through the "top", $\|q\| = \text{const}$, of ∂P.

We will show that

(a) the flow of $\widetilde{Q}$ is equal to the flow of Q on P and on the disk $q = 0$, $\|p\| \le b_0$,

(b) $\pounds_{X_{\widetilde{Q}}}\|q\|^2 \ge 0$ when $\|p\| \le b_0$, with strict inequality when $q \ne 0$,

(c) $\pounds_{X_{\widetilde{Q}}}\|p\|^2 < 0$ when $\|p\| \ge b_0$.

It is not hard to see that (iv) readily follows from these assertions.

Assertion (a) is obvious since $\widetilde{Q} = Q$ in the region where $|Q| \le c$ and $\|q\| \le a_0$ and $\|p\| \le b_0$, containing $V \supset P$ and the disk $q = 0$, $\|p\| \le b_0$. It remains to check (b) and (c), that is, that the function $\|q\|^2$ increases along the flow of $\widetilde{Q}$ when $\|p\| \le b_0$, and the function $\|p\|^2$ decreases along the flow of $\widetilde{Q}$ when $\|p\| \ge b_0$, unless $q = 0$.

To prove (b), first note that $\widetilde{Q} = \widehat{Q}$ in the region where $\|p\| \le b_0$. Also recall that $\eta' \ge \epsilon/2$, and hence

$$\epsilon\phi(\|q\|) + \big(1 - \phi(\|q\|)\big)(\eta' \circ Q) \ge \epsilon/2.$$

Therefore since $\phi(\|q\|)$ is independent of p, we have

$$\begin{aligned}\pounds_{X_{\widetilde{Q}}}\|q\|^2 = \pounds_{X_{\widehat{Q}}}\|q\|^2 &= \epsilon\phi(\|p\|)\pounds_{X_Q}\|q\|^2 + \big(1 - \phi(\|p\|)\big)(\eta' \circ Q)\pounds_{X_Q}\|q\|^2 \\ &\ge \frac{\epsilon\lambda}{2}\|q\|^2 \ge 0.\end{aligned}$$

Let us now establish (c), which is somewhat more involved than (b) due to the q-dependence of ϕ. Writing ψ for $\psi(\|p\|)$ and ϕ for $\phi(\|q\|)$ and ϕ' for $\phi'(\|q\|)$, we have

$$\begin{aligned}\pounds_{X_{\widetilde{Q}}}\|p\|^2 &= \epsilon\psi\pounds_{X_Q}\|p\|^2 + (1-\psi)\pounds_{X_{\widehat{Q}}}\|p\|^2 \\ &= \epsilon\psi\pounds_{X_Q}\|p\|^2 + (1-\psi)\big(\epsilon\phi\pounds_{X_Q}\|p\|^2 + (1-\phi)(\eta' \circ Q)\pounds_{X_Q}\|p\|^2\big) \\ &\qquad + (1-\psi)(\epsilon Q - \eta \circ Q)\phi'\pounds_{X_{\|q\|}}\|p\|^2 \\ &= \big[\epsilon\psi + (1-\psi)\big(\epsilon\phi + (1-\phi)(\eta' \circ Q)\big)\big]\pounds_{X_Q}\|p\|^2 \\ &\qquad + (1-\psi)(\epsilon Q - \eta \circ Q)\phi'\pounds_{X_{\|q\|}}\|p\|^2.\end{aligned}$$

To bound from above the first term in this expression, note that the coefficient of $\pounds_{X_Q}\|p\|^2$ is positive and satisfies

$$\epsilon\psi + (1-\psi)\big(\epsilon\phi + (1-\phi)(\eta' \circ Q)\big) \ge \epsilon\psi + \frac{\epsilon}{2}(1-\psi) = \frac{\epsilon}{2}(1+\psi) \ge \frac{\epsilon}{2}.$$

Therefore

$$\big[\epsilon\psi + (1-\psi)\big(\epsilon\phi + (1-\phi)(\eta' \circ Q)\big)\big]£_{X_Q}\|p\|^2 \le -\epsilon\frac{\lambda}{2}\|p\|^2.$$

Bounding from above the second term, our choices of $a_0 = r/\sqrt{\epsilon}$ and $a_1 = 2a_0$ and also the requirement that $|\phi'| \le 2/|a_1 - a_0| = 2/a_0$ enter the picture. Then by (4-5), we have

$$2|\epsilon Q - \eta \circ Q|\,|\phi'|\,\|p\| \le 8c\frac{2}{|a_1 - a_0|}\|p\| = \frac{16c}{a_0}\|p\| \le \frac{16c\sqrt{\epsilon}}{r}\|p\|.$$

Thus in the region where $\|p\| > b_0 = 32c/\lambda r\sqrt{\epsilon}$, we have

$$£_{X_{\tilde{Q}}}\|p\|^2 < -\epsilon\frac{\lambda}{2}\|p\|^2 + \frac{16c\sqrt{\epsilon}}{r}\|p\| < 0,$$

completing the proof of (c). (Note that b_0 is chosen exactly to make (c) hold.) □

Proof of Theorem 1.4. As pointed out in the introduction, we have a more general result in dimension two:

Theorem 4.5. *Let $H : S^1 \times \mathbb{R}^2 \to \mathbb{R}$ be a Hamiltonian which is equal to a hyperbolic quadratic form at infinity. Assume that φ_H has at least two isolated homologically nontrivial fixed points and* $\mathrm{Fix}(\varphi_H)$ *is finite. Then φ_H has simple periodic orbits of arbitrarily large period.*

Remark 4.6. A similar two-dimensional result holds when H is elliptic quadratic at infinity. (In fact, the requirement that the fixed points be homologically nontrivial is not needed in this case.) Indeed since the Hamiltonian is elliptic outside a compact set, a sufficiently large sublevel will be invariant under the flow. Then Franks's theorem [1992] stating that an area-preserving map of the two-disk has either one or infinitely many periodic points implies the result.

Remark 4.7. Using Theorem 4.5, we can also prove a weaker version of Franks's theorem on S^2, asserting that a Hamiltonian diffeomorphism of S^2 with a hyperbolic fixed point must necessarily have infinitely many periodic orbits. However the argument is somewhat involved, and we omit it since this result also follows from the main theorem of [Ginzburg and Gürel 2014], and as has been mentioned in the introduction, at least two other symplectic proofs of Franks's theorem are available; see [Bramham and Hofer 2012; Collier et al. 2012; Kerman 2012].

Proof of Theorem 4.5. Observe that if φ_H has a homologically nontrivial fixed point with nonzero mean index, then the theorem follows from Theorem 4.1. So let us assume that there are at least two isolated homologically nontrivial fixed points with zero mean index. Notice that all of these points cannot have nonzero local Floer homology concentrated in degree zero: $\mathrm{HF}_*(H) = 0$ when $* \ne 0$ and

$\mathrm{HF}_0(H) = \mathbb{Z}_2$. Thus φ_H must have at least one fixed point with zero mean index and nonzero local Floer homology in degree ± 1. Such an orbit is a symplectically degenerate maximum, and its presence implies that φ_H has simple periodic orbits of arbitrarily large prime period; see [Ginzburg and Gürel 2009a; 2010] and also [Ginzburg 2010; Hein 2012; Hingston 2009]. (Strictly speaking, the latter fact has been established only for (a broad class of) closed symplectic manifolds. However since the Hamiltonian in our case is a compactly supported perturbation of a hyperbolic quadratic form on $\mathbb{R}^{2n}$, having periodic orbits only within the support of the perturbation, the proof in the case of closed manifolds, for instance, the one in [Ginzburg and Gürel 2009a], goes through word for word.) $\square$

Proof of Theorem 1.4 in dimension four. Recall that the homology is concentrated in degree zero and $\mathrm{HF}_0(H) = \mathbb{Z}_2$. Hence one of the fixed points of φ_H must have nonzero Conley–Zehnder index. By a straightforward index analysis, it is easy to see that in dimension four such an orbit must necessarily have nonzero mean index. (Indeed observe that in dimension four the mean index of a nondegenerate fixed point is zero if and only if the linearization is hyperbolic or its eigenvalues comprise two pairs "conjugate" to each other. It is clear that in both cases the Conley–Zehnder index is zero.) Finally applying Theorem 4.1, we obtain the existence of simple periodic orbits with arbitrarily large prime period. $\square$

Acknowledgements. The author is grateful to Alberto Abbondandolo, Viktor Ginzburg, Leonid Polterovich and Cem Yalçın Yıldırım for useful discussions, comments and suggestions and to the referee for valuable remarks.

References

[Abbondandolo 2001] A. Abbondandolo, *Morse theory for Hamiltonian systems*, Research Notes in Mathematics **425**, Chapman & Hall/CRC, Boca Raton, FL, 2001. MR 2002e:37103 Zbl 0967.37002

[Amann and Zehnder 1980] H. Amann and E. Zehnder, "Periodic solutions of asymptotically linear Hamiltonian systems", *Manuscripta Math.* **32**:1-2 (1980), 149–189. MR 82i:58026 Zbl 0443.70019

[Antonacci 1997] F. Antonacci, "Existence of periodic solutions of Hamiltonian systems with potential indefinite in sign", *Nonlinear Anal.* **29**:12 (1997), 1353–1364. MR 98k:34067 Zbl 0894.34036

[Arnold 1978] V. I. Arnold, *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics **60**, Springer, New York, 1978. MR 57 #14033b Zbl 0386.70001

[Baker et al. 2001] R. C. Baker, G. Harman, and J. Pintz, "The difference between consecutive primes, II", *Proc. London Math. Soc.* (3) **83**:3 (2001), 532–562. MR 2002f:11125 Zbl 1016.11037

[Batoreo 2013] M. Batoreo, "On hyperbolic points and periodic orbits of symplectomorphisms", preprint, 2013. arXiv 1310.1974

[Bramham and Hofer 2012] B. Bramham and H. Hofer, "First steps towards a symplectic dynamics", pp. 127–178 in *Algebra and geometry: In memory of C. C. Hsiung*, edited by H.-D. Cao and S.-T. Yau, Surveys in Differential Geometry **17**, International Press, Boston, 2012.

[Chance et al. 2013] M. Chance, V. Ginzburg, and B. Gürel, "Action-index relations for perfect Hamiltonian diffeomorphisms", *J. Symplectic Geom.* **11**:3 (2013), 449–474. MR 3100801 Zbl 06221957

[Cieliebak et al. 1995] K. Cieliebak, A. Floer, and H. Hofer, "Symplectic homology, II: A general construction", *Math. Z.* **218**:1 (1995), 103–122. MR 95m:58055 Zbl 0869.58011

[Collier et al. 2012] B. Collier, E. Kerman, B. M. Reiniger, B. Turmunkh, and A. Zimmer, "A symplectic proof of a theorem of Franks", *Compos. Math.* **148**:6 (2012), 1969–1984. MR 2999311 Zbl 1267.53093

[Cornea 2001] O. Cornea, "Homotopical dynamics, III: Real singularities and Hamiltonian flows", *Duke Math. J.* **109**:1 (2001), 183–204. MR 2002j:37024 Zbl 1107.37300

[Demmel 1997] J. W. Demmel, *Applied numerical linear algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. MR 98m:65001 Zbl 0879.65017

[Entov and Polterovich 2009] M. Entov and L. Polterovich, "Rigid subsets of symplectic manifolds", *Compos. Math.* **145**:3 (2009), 773–826. MR 2011a:53174 Zbl 1230.53080

[Floer 1989a] A. Floer, "Symplectic fixed points and holomorphic spheres", *Comm. Math. Phys.* **120**:4 (1989), 575–611. MR 90e:58047 Zbl 0755.58022

[Floer 1989b] A. Floer, "Witten's complex and infinite-dimensional Morse theory", *J. Differential Geom.* **30**:1 (1989), 207–221. MR 90d:58029 Zbl 0678.58012

[Franks 1992] J. Franks, "Geodesics on S^2 and periodic points of annulus homeomorphisms", *Invent. Math.* **108**:2 (1992), 403–418. MR 93f:58192 Zbl 0766.53037

[Franks 1996] J. Franks, "Area preserving homeomorphisms of open surfaces of genus zero", *New York J. Math.* **2** (1996), 1–19. MR 97c:58123 Zbl 0891.58033

[Franks and Handel 2003] J. Franks and M. Handel, "Periodic points of Hamiltonian surface diffeomorphisms", *Geom. Topol.* **7** (2003), 713–756. MR 2004j:37101 Zbl 1034.37028

[Ginzburg 2007] V. L. Ginzburg, "Coisotropic intersections", *Duke Math. J.* **140**:1 (2007), 111–163. MR 2009h:53200 Zbl 1129.53062

[Ginzburg 2010] V. L. Ginzburg, "The Conley conjecture", *Ann. of Math.* (2) **172**:2 (2010), 1127–1180. MR 2011h:53127 Zbl 1228.53098

[Ginzburg and Gürel 2009a] V. L. Ginzburg and B. Z. Gürel, "Action and index spectra and periodic orbits in Hamiltonian dynamics", *Geom. Topol.* **13**:5 (2009), 2745–2805. MR 2010k:53149 Zbl 1172.53052

[Ginzburg and Gürel 2009b] V. L. Ginzburg and B. Z. Gürel, "On the generic existence of periodic orbits in Hamiltonian dynamics", *J. Mod. Dyn.* **3**:4 (2009), 595–610. MR 2011b:53210 Zbl 1190.53084

[Ginzburg and Gürel 2010] V. L. Ginzburg and B. Z. Gürel, "Local Floer homology and the action gap", *J. Symplectic Geom.* **8**:3 (2010), 323–357. MR 2011g:53192 Zbl 1206.53087

[Ginzburg and Gürel 2012] V. L. Ginzburg and B. Z. Gürel, "Conley conjecture for negative monotone symplectic manifolds", *Int. Math. Res. Not.* **2012**:8 (2012), 1748–1767. MR 2920829 Zbl 1242.53100

[Ginzburg and Gürel 2014] V. L. Ginzburg and B. Z. Gürel, "Hyperbolic fixed points and periodic orbits of Hamiltonian diffeomorphisms", *Duke Math. J.* **163**:3 (2014), 565–590. MR 3165423

[Gürel 2013] B. Z. Gürel, "On non-contractible periodic orbits of Hamiltonian diffeomorphisms", *Bull. Lond. Math. Soc.* **45**:6 (2013), 1227–1234. Zbl 06237636

[Hein 2012] D. Hein, "The Conley conjecture for irrational symplectic manifolds", *J. Symplectic Geom.* **10**:2 (2012), 183–202. MR 2926994 Zbl 1275.37026

[Hingston 2009] N. Hingston, "Subharmonic solutions of Hamiltonian equations on tori", *Ann. of Math.* (2) **170**:2 (2009), 529–560. MR 2010j:53178 Zbl 1180.58009

[Hofer and Zehnder 1994] H. Hofer and E. Zehnder, *Symplectic invariants and Hamiltonian dynamics*, Birkhäuser, Basel, 1994. MR 96g:58001 Zbl 0805.58003

[Kerman 2012] E. Kerman, "On primes and period growth for Hamiltonian diffeomorphisms", *J. Mod. Dyn.* **6**:1 (2012), 41–58. MR 2929129 Zbl 1244.53095

[Le Calvez 2006] P. Le Calvez, "Periodic orbits of Hamiltonian homeomorphisms of surfaces", *Duke Math. J.* **133**:1 (2006), 125–184. MR 2007a:37052 Zbl 1101.37031

[Long 2002] Y. Long, *Index theory for symplectic paths with applications*, Progress in Mathematics **207**, Birkhäuser, Basel, 2002. MR 2003d:37091 Zbl 1012.37012

[Mawhin and Willem 1989] J. Mawhin and M. Willem, *Critical point theory and Hamiltonian systems*, Applied Mathematical Sciences **74**, Springer, New York, 1989. MR 90e:58016 Zbl 0676.58017

[McDuff 1991] D. McDuff, "Symplectic manifolds with contact type boundaries", *Invent. Math.* **103**:3 (1991), 651–671. MR 92e:53042 Zbl 0719.53015

[McDuff and Salamon 2004] D. McDuff and D. Salamon, *J-holomorphic curves and symplectic topology*, American Mathematical Society Colloquium Publications **52**, Amer. Math. Soc., Providence, RI, 2004. MR 2004m:53154 Zbl 1064.53051

[Rabinowitz 1980] P. H. Rabinowitz, "On subharmonic solutions of Hamiltonian systems", *Comm. Pure Appl. Math.* **33**:5 (1980), 609–633. MR 81k:34032 Zbl 0425.34024

[Salamon 1999] D. Salamon, "Lectures on Floer homology", pp. 143–229 in *Symplectic geometry and topology* (Park City, UT, 1997), edited by Y. Eliashberg and L. Traynor, IAS/Park City Math. Ser. **7**, Amer. Math. Soc., Providence, RI, 1999. MR 2000g:53100 Zbl 1031.53118

[Salamon and Zehnder 1992] D. Salamon and E. Zehnder, "Morse theory for periodic solutions of Hamiltonian systems and the Maslov index", *Comm. Pure Appl. Math.* **45**:10 (1992), 1303–1360. MR 93g:58028 Zbl 0766.58023

[Schwarz 2000] M. Schwarz, "On the action spectrum for closed symplectically aspherical manifolds", *Pacific J. Math.* **193**:2 (2000), 419–461. MR 2001c:53113 Zbl 1023.57020

[Seidel 2008] P. Seidel, "A biased view of symplectic cohomology", pp. 211–253 in *Current developments in mathematics, 2006*, edited by B. Mazur et al., Int. Press, Somerville, MA, 2008. MR 2010k:53153 Zbl 1165.57020

[Viterbo 1999] C. Viterbo, "Functors and computations in Floer homology with applications, I", *Geom. Funct. Anal.* **9**:5 (1999), 985–1033. MR 2000j:53115 Zbl 0954.57015

[Williamson 1936] J. Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems", *Amer. J. Math.* **58**:1 (1936), 141–163. MR 1507138 Zbl 0013.28401

[Zhang and Liu 2011] Q. Zhang and C. Liu, "Infinitely many periodic solutions for second order Hamiltonian systems", *J. Differential Equations* **251**:4-5 (2011), 816–833. MR 2012e:34095 Zbl 1230.37081

[Zou 2001] W. Zou, "Multiple solutions for second-order Hamiltonian systems via computation of the critical groups", *Nonlinear Anal.* **44**:7, Ser. A: Theory Methods (2001), 975–989. MR 2002c:37100 Zbl 0997.37039

Received May 27, 2013. Revised November 28, 2013.

BAŞAK Z. GÜREL
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CENTRAL FLORIDA
ORLANDO, FL 32816
UNITED STATES
basak.gurel@ucf.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.183

NONSPLITTABILITY OF THE RATIONAL HOMOLOGY COBORDISM GROUP OF 3-MANIFOLDS

SE-GOO KIM AND CHARLES LIVINGSTON

Let $\mathbb{Z}[1/p]$ denote the ring of integers with the prime p inverted. There is a canonical homomorphism $\Psi : \bigoplus \Theta^3_{\mathbb{Z}[1/p]} \to \Theta^3_{\mathbb{Q}}$, where Θ^3_R denotes the three-dimensional smooth R-homology cobordism group of R-homology spheres and the direct sum is over all prime integers. Gauge-theoretic methods prove the kernel is infinitely generated. Here we prove that Ψ is not surjective, with cokernel infinitely generated. As a basic example we show that for p and q distinct primes, there is no rational homology cobordism from the lens space $L(pq, 1)$ to any $M_p \# M_q$, where $H_1(M_p) = \mathbb{Z}_p$ and $H_1(M_q) = \mathbb{Z}_q$. More subtle examples include cases in which a cobordism to such a connected sum exists topologically but not smoothly. (Conjecturally such a splitting always exists topologically.) Further examples can be chosen to represent 2-torsion in $\Theta^3_{\mathbb{Q}}$. Let $\mathcal{K}$ denote the kernel of $\Theta^3_{\mathbb{Q}} \to \widehat{\Theta}^3_{\mathbb{Q}}$, where $\widehat{\Theta}^3_{\mathbb{Q}}$ denotes the topological homology cobordism group. Freedman proved that $\Theta^3_{\mathbb{Z}} \subset \mathcal{K}$. A corollary of results here is that $\mathcal{K}/\Theta^3_{\mathbb{Z}}$ is infinitely generated. We also demonstrate the failure in dimension three of splitting theorems that apply to higher-dimensional knot concordance groups.

1. Introduction

Furuta [1990] applied instanton gauge theory to reveal unexpectedly deep structure in the homology cobordism group of smooth homology 3-spheres, $\Theta^3_{\mathbb{Z}}$. Here we will use the added algebraic structures associated to Heegaard–Floer theory to identify further complications in the rational cobordism group, $\Theta^3_{\mathbb{Q}}$.

As a simple example, an application of the rational homology cobordism classification of lens spaces [Lisca 2007] implies that for p and q relatively prime, the lens space $L(pq, 1)$ is not $\mathbb{Q}$-homology cobordant to any connected sum $L(p, a) \# L(q, b)$. A simple consequence of the work here is that $L(pq, 1)$ is not $\mathbb{Q}$-homology cobordant to any connected sum $M_p \# M_q$ where $H_1(M_p) = \mathbb{Z}_p$

This work was supported in part by the National Science Foundation under Grant 1007196 and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) NRF-2011-0012893.
MSC2010: primary 57M27; secondary 57M25.
Keywords: three-manifold, connected sum, homology cobordism, knot concordance.

and $H_1(M_q) = \mathbb{Z}_q$. We let Θ^3_R denote the R-homology cobordism group of three-dimensional R-homology spheres. Note that $\Theta^3_{\mathbb{Z}[1/p]}$ is generated by three-manifolds M with $H_1(M)$ p-torsion. There is a canonical map

$$\Phi : \bigoplus_{p\in\mathcal{P}} \Theta^3_{\mathbb{Z}[1/p]} \to \Theta^3_{\mathbb{Q}}.$$

Rochlin's theorem and Furuta's result imply that the kernel of Φ is infinitely generated. To simplify notation, we abbreviate the summation by $\bigoplus \Theta_p$; our main result is the following:

Proposition. *The cokernel of* Φ, $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$, *contains a free subgroup of infinite rank generated by lens spaces of the form* $L(pq, 1)$. *It also contains an infinite subgroup generated by elements of order two*: *lens spaces of the form* $L(4n^2 + 1, 2n)$. *An infinite subgroup is also generated by three-manifolds that bound* $\mathbb{Q}$*-homology balls topologically.*

We also present applications to the study of knot concordance and present families of elements in the kernel $\Theta^3_{\mathbb{Q}}/\Theta^3_{\mathbb{Z}} \to \widehat{\Theta}^3_{\mathbb{Q}}$, where $\widehat{\Theta}^3_{\mathbb{Q}}$ denotes the topological cobordism group. Similar examples were presented in [Hedden et al. 2012], with the additional condition that bordisms were assumed to be spin.

An important perspective is provided by considering the torsion linking form of three-manifolds, which yields a homomorphism $\Theta^3_{\mathbb{Q}} \to W(\mathbb{Q}/\mathbb{Z})$, the Witt group of nondegenerate symmetric $\mathbb{Q}/\mathbb{Z}$-valued bilinear forms on finite abelian groups. According to [Kawauchi and Kojima 1980] this homomorphism is surjective. Again by Rochlin's theorem and Furuta's result, it has infinitely generated kernel (in the topological category it is conjecturally an isomorphism). There is a canonical isomorphism $\bigoplus_{p\in\mathcal{P}} W(\mathbb{F}_p) \to W(\mathbb{Q}/\mathbb{Z})$, where $W(\mathbb{F}_p)$ is the Witt group of nondegenerate symmetry bilinear forms on $\mathbb{F}_p$-vector spaces and $\mathcal{P}$ is the set of prime integers. The conjecture that topological cobordism is determined by the linking form implies that $\widehat{\Theta}^3_{\mathbb{Q}}$ has a corresponding primary decomposition. One thrust of our work here is to display the extent of the failure of the existence of such a primary decomposition in the smooth setting.

The commutative diagram in Figure 1 organizes the groups of interest.

The proposition above states that $\Theta^3_{\mathbb{Q}}/\mathrm{Image}(\Phi)$ is infinitely generated containing an infinite free subgroup and infinite two-torsion and that, furthermore, the image of $\mathcal{K}$ in $\Theta^3_{\mathbb{Q}}/\mathrm{Image}(\Phi)$ similarly contains an infinite subgroup.

Definition. A three-manifold M is said to *split* if it represents a class in the image of Φ. That is, a manifold does not split if it is nontrivial in the cokernel of Φ.

Outline. In Sections 2, 3, and 4 we present some of the basic definitions used throughout the paper, isolate a basic result concerning metabolizers of nondegenerate

$$
\begin{tikzcd}
 & \bigoplus \Theta_p \arrow[r] \arrow[d, "\Phi"] & \bigoplus \widehat{\Theta}_p \arrow[r] \arrow[d, "\widehat{\Phi}"] & \bigoplus_{p \in \mathcal{P}} W(\mathbb{F}_p) \arrow[d, "\cong"] \\
\mathcal{K} \arrow[r] & \Theta^3_{\mathbb{Q}} \arrow[r] & \widehat{\Theta}^3_{\mathbb{Q}} \arrow[r] & W(\mathbb{Q}/\mathbb{Z})
\end{tikzcd}
$$

Figure 1. Groups of interest. Hats denote the topological category and $\mathcal{K}$ denotes the kernel of the canonical homomorphism from the smooth to the topological $\mathbb{Q}$-homology cobordism group. With the exception of the inclusion of the kernel, all horizontal arrows are surjective. Conjecturally the right square consists of isomorphisms.

symmetric bilinear forms, and discuss Spin^c structures. Section 5 presents one of our main results, describing an obstruction based on Heegaard–Floer d-invariants to a class in $\Theta^3_{\mathbb{Q}}$ being in the image $\Phi(\bigoplus \Theta_p)$. Following this we provide a series of examples:

- Section 6 demonstrates that lens spaces $L(pq, 1)$ with p and q square-free and relatively prime do not split, and extends this to finite connected sums of such lens spaces, with all p and q distinct, thus proving that $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$ is infinite. Section 7 further extends this, demonstrating that the set of lens spaces of the form $L(pq, 1)$ (with p and q now required to be prime) generate an infinite free subgroup of infinite rank contained in $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$.
- Section 8 considers specific lens spaces of the form $L(4n^2 + 1, 2n)$ to provide elements of order 2 in $\Theta^3_{\mathbb{Q}}$ that do not split, in particular showing that $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$ contains 2-torsion. Section 9 expands on this, providing an infinite family of independent elements of order 2.
- Section 10 begins the examination of the failure of splittings among manifolds that do split topologically; that is, we consider manifolds representing classes in $\mathcal{K}$. The main example is built from surgery on the connected sum of the torus knot $T_{3,5}$ and the untwisted Whitehead double of the trefoil knot, $\mathrm{Wh}(T_{2,3}) = D$. We show that $S^3_{15}(T_{3,5} \# D)$ splits topologically but not smoothly. Section 11 generalizes that example to an infinite family, using $(p, p+2)$ torus knots, with p odd.
- Section 12 applies the results of Section 6 to demonstrate the failure of a splitting theorem for knot concordance which, by a result of Stoltzfus [1977], applies algebraically and in dimensions greater than 3.
- According to [Freedman 1982; Freedman and Quinn 1990], all homology spheres bound contractible 4-manifolds topologically, so $\Theta^3_{\mathbb{Z}} \subset \mathcal{K}$. In Section 13

we outline the proof that the quotient $\mathcal{H}/\Theta^3_{\mathbb{Z}}$ contains an infinitely generated free subgroup. This was proved in [Hedden et al. 2012] with the added constraint that one restricts the cobordism groups by considering only manifolds that are $\mathbb{Z}_2$-homology spheres or by requiring that all spaces have spin structures. We briefly indicate how results here permit one to remove those restrictions in the argument in the same reference.

2. Definitions

We will consider $\mathbb{Q}$-homology 3-spheres: these are closed 3-manifolds M^3 with $H_n(M^3, \mathbb{Q}) \cong H_n(S^3, \mathbb{Q})$ for all n. For each such M there is a symmetric linking form $\beta : H_1(M) \times H_1(M) \to \mathbb{Q}/\mathbb{Z}$ which is nondegenerate in the sense that the induced map $\beta^* : H_1(M) \to \mathrm{Hom}(H_1(M), \mathbb{Q}/\mathbb{Z})$ is an isomorphism. If $M = \partial X^4$, where X is a compact 4-manifold and $H_n(X, \mathbb{Q}) = H_n(B^4, \mathbb{Q})$ for all n, then the kernel $\mathcal{M}$ of the map $H_1(M) \to H_1(X)$ is a metabolizer for β (see [Casson and Gordon 1986]). That is, $\mathcal{M}^\perp = \mathcal{M}$, and in particular $|\mathcal{M}|^2 = |H_1(M)|$. The Witt group $W(\mathbb{Q}/\mathbb{Z})$ is built from the set of all pairs (G, β) where G is a finite abelian group and β is a nondegenerate symmetric bilinear form taking values in $\mathbb{Q}/\mathbb{Z}$. There is an equivalence relation on this set: $(G, \beta) \sim (G', \beta')$ if $(G \oplus G', \beta \oplus -\beta')$ has a metabolizer, and under this relation it becomes an abelian group under direct sum, denoted $W(\mathbb{Q}/\mathbb{Z})$. It can be proved (for instance, see [Alexander et al. 1976]) that a pair (G, β) is Witt trivial if and only if it has a metabolizer. The proof of this fact includes the following, which we will be using.

Proposition 1. *If $(G_1, \beta_1) \oplus (G_2, \beta_2)$ has metabolizer $\mathcal{M}$ and (G_2, β_2) has metabolizer $\mathcal{M}_2$, then $\mathcal{M}_1 = \{g \in G_1 \mid (g, h) \in \mathcal{M}$ for some $h \in \mathcal{M}_2\}$ is a metabolizer for (G_1, β_1).*

The Witt groups $W(\mathbb{Q}/\mathbb{Z}, \langle p \rangle)$ are defined as is $W(\mathbb{Q}/\mathbb{Z})$, considering only p-torsion abelian groups, and the decomposition $W(\mathbb{Q}/\mathbb{Z}) \cong \bigoplus_{p \in \mathcal{P}} W(\mathbb{Q}/\mathbb{Z}, \langle p \rangle)$ is easily proved. The Witt group of nondegenerate symmetric forms on $\mathbb{F}_p$-vector spaces is denoted $W(\mathbb{F}_p)$. The inclusion $W(\mathbb{F}_p) \to W(\mathbb{Q}/\mathbb{Z}, \langle p \rangle)$ is an isomorphism. In the proof of this, the inclusion is clearly injective, and an inverse map $W(\mathbb{Q}/\mathbb{Z}, \langle p \rangle) \to W(\mathbb{F}_p)$ is explicitly constructed via "devissage" [Alexander et al. 1976; Milnor and Husemoller 1973].

Let R be a commutative ring. Two closed 3-manifolds, M_1 and M_2, are called *R-homology cobordant* if there is a compact smooth 4-manifold X with boundary the disjoint union $M_1 \cup -M_2$ such that the inclusions $H_*(M_i, R) \to H_*(X, R)$ are isomorphisms. Equivalently they are R-cobordant, written $M_1 \sim_R M_2$, if $M_1 \# -M_2$ bounds an R-homology 4-ball. The set of R-cobordism classes of R-homology spheres forms an abelian group with operation induced by connected sum. This group is denoted Θ^3_R.

The ring $\mathbb{Z}[1/p]$ is the ring of integers with p inverted, consisting of all rational numbers with denominator a power of p. A closed 3-manifold M is a $\mathbb{Z}[1/p]$-homology sphere if and only if $H_1(M)$ is p-torsion. The linking form provides well-defined homomorphisms $\Theta^3_{\mathbb{Q}} \to W(\mathbb{Q}/\mathbb{Z})$ and $\Theta^3_{\mathbb{Z}[1/p]} \to W(\mathbb{F}_p)$ for which the following diagram commutes. As in the introduction, we abbreviate $\Theta^3_{\mathbb{Z}[1/p]}$ by Θ_p.

$$\begin{tikzcd} \bigoplus \Theta_p \arrow[r] \arrow[d, "\Phi"] & \bigoplus_{p\in\mathcal{P}} W(\mathbb{F}_p) \arrow[d, "\cong"] \\ \Theta^3_{\mathbb{Q}} \arrow[r] & W(\mathbb{Q}/\mathbb{Z}) \end{tikzcd}$$

If we switch to the topological category, all these maps are conjecturally isomorphisms.

3. Metabolizers for connected sums

3.1. ***Metabolizers.*** If a connected sum of 3-manifolds bounds a rational homology ball, the associated metabolizer of the linking form does not necessarily split relative to the connected sum. As a simple example, for the connected sum of lens spaces $L(p,1)\,\#-L(p,1)$ with p prime, the only metabolizers for the linking form on $\mathbb{Z}_p \oplus \mathbb{Z}_p$ are the diagonal and skew diagonal subgroups, generated by $(1,\pm 1)$. However the existence of the connected sum decomposition does place constraints on the metabolizer.

Theorem 2. *If p is prime, G is a finite abelian group, and a given nondegenerate symmetric bilinear form $\beta_1 \oplus \beta_2$ on $\mathbb{Z}_p \oplus G$ has metabolizer $\mathcal{M}$, then for some $a \in G$, $(1,a) \in \mathcal{M}$.*

Proof. Let G_p denote the p-torsion in G. There is a metabolizer $\mathcal{M}_p$ for the form restricted to $\mathbb{Z}_p \oplus G_p$. If $\mathcal{M}_p \subset G_p$, then it would represent a metabolizer for the linking form restricted to G_p, implying that the order of G_p is an even power of p. But since the form on $\mathbb{Z}_p \oplus G_p$ is metabolic, the order of G_p must be an odd power of p. It follows that there is an element $(a',a'') \in \mathcal{M}_p$ with $a' \neq 0$. Multiplying by $(a')^{-1} \mod p$, we see that $(1,a) \in \mathcal{M}_p \subset \mathcal{M}$ for some $a \in G_p$. □

In the following corollary, for each integer k, G_k denotes a finite abelian group of order dividing a power of k.

Corollary 3. *If m is a square-free integer, $G_m \oplus G_n$ is a finite abelian group with* $\gcd(m,n) = 1$ *and a given nondegenerate symmetric bilinear form $\beta_1 \oplus \beta_2 \oplus \beta_3$ on $\mathbb{Z}_m \oplus G_m \oplus G_n$ has metabolizer $\mathcal{M}$, then for some $a \in G_m$, $(1,a,0) \in \mathcal{M}$.*

Proof. Write $\mathbb{Z}_m = \mathbb{Z}_{p_1} \oplus \cdots \oplus \mathbb{Z}_{p_k}$ with each p_i prime. (Since m is square-free, there are no prime powers in the factorization of m and all p_i are distinct.) By Theorem 2, the projection of $\mathcal{M}$ to each $\mathbb{Z}_{p_i}$ summand is surjective. Since the p_i are relatively prime, the projection to $\mathbb{Z}_m$ is similarly surjective. □

In order to construct elements of infinite order, we will need to consider multiples of linking forms. Without loss of generality, we will be able to assume that the multiplicative factors are divisible by four.

Theorem 4. *Suppose that p is prime and a nondegenerate symmetric bilinear form $4k(\beta_1 \oplus \beta_2)$ on $(\mathbb{Z}_p \oplus G)^{4k}$ has a metabolizer $\mathcal{M}$. Then $\mathcal{M}$ contains an element of the form $(1, 1, \ldots, 1, \alpha_{2k+1}, \ldots, \alpha_{4k}) \oplus b$ for some set of $\alpha_i \in \mathbb{Z}_p$ and some $b \in G^{4k}$.*

Proof. The Witt group $W(\mathbb{Q}/\mathbb{Z})$ is 4-torsion [Milnor and Husemoller 1973], and thus $4k\beta_2$ has a metabolizer $\mathcal{M}'$. By Proposition 1, the set of elements x such that $(x, y) \in \mathcal{M}$ for some $y \in \mathcal{M}'$ is a metabolizer, denoted $\mathcal{N}$, for $4k\beta_1$, and thus is $2k$-dimensional. As argued in [Livingston and Naik 1999], a simple application of the Gauss–Jordan algorithm applied to a generating set for $\mathcal{N}$ yields a generating set consisting of vectors of the form $(1, 0, 0, 0, \ldots, 0, *, *, \ldots)$, $(0, 1, 0, 0, \ldots, 0, *, *, \ldots)$, $(0, 0, 1, 0, \ldots, 0, *, *, \ldots)$, …, where each initial sequence of a 1 and 0s is of length $2k$.

By adding these vectors together, we find that the metabolizer $\mathcal{N}$ contains an element of the form $(1, 1, \ldots, 1, \alpha_{2k+1}, \ldots, \alpha_{4k}) \in \mathbb{Z}_p^{4k}$. Finally, since each element in $\mathcal{N}$ pairs with an element in the metabolizer $\mathcal{M}'$ to give an element in $\mathcal{M}$, we get the desired element b. □

4. Spin^c structures

We need the following facts about $\mathrm{Spin}^c(Y)$, the set of Spin^c structures on manifolds.

- The first Chern class is a map $c_1 : \mathrm{Spin}^c(Y) \to H^2(Y)$.
- There is a transitive action $H^2(Y) \times \mathrm{Spin}^c(Y) \to \mathrm{Spin}^c(Y)$ denoted $(\alpha, \mathfrak{s}) \to \alpha \cdot \mathfrak{s}$.
- For $Y \subset W$ a codimension-one submanifold with trivial normal bundle, such as a boundary component of W, the restriction map r is functorial: If $\mathfrak{s} \in \mathrm{Spin}^c(W)$, $\alpha \in H^2(W)$, then
$$r(\alpha \cdot \mathfrak{s}) = r(\alpha) \cdot r(\mathfrak{s}).$$
- For all $\alpha \in H^2(Y)$ and $\mathfrak{s} \in \mathrm{Spin}^c(Y)$, $c_1(\alpha \cdot \mathfrak{s}) - c_1(\mathfrak{s}) = 2\alpha$.
- As a corollary, if $|H^2(Y)|$ is finite and odd, then $c_1 : \mathrm{Spin}^c(Y) \to H^2(Y)$ is a bijection.
- There is a canonical bijection $\mathrm{Spin}^c(Y_1 \# Y_2) \to \mathrm{Spin}^c(Y_1) \times \mathrm{Spin}^c(Y_2)$.

For every smooth 4-manifold W, the set $\mathrm{Spin}^c(W)$ is nonempty. (See [Gompf and Stipsicz 1999] for a proof.) As a consequence, we have the following.

Theorem 5. *Let $N = \partial X$ and let $\mathfrak{s} \in \mathrm{Spin}^c(N)$ be the restriction of a Spin^c structure on X. Then the Spin^c structures on N which extend to X are those of the form $\alpha \cdot \mathfrak{s}$ for α in the image of the restriction map $r : H^2(X) \to H^2(N)$.*

4.1. ***Identifying $H_1(N)$ and $H^2(N)$.*** Let N be a rational homology 3-sphere bounding a rational homology ball X. Then, by Poincaré duality, $H_1(N) \cong H^2(N)$. We have denoted the kernel of $H_1(N) \to H_1(X)$ by $\mathcal{M}$. Via duality, it corresponds to the image of $H^2(X)$ in $H^2(N)$. Thus we will use $\mathcal{M}$ to denote this subgroup of $H^2(N)$.

4.2. ***Spin structures.*** If the order $|H_1(M)|$ is odd, then there is a unique spin structure on M that lifts to a canonical Spin^c structure that we will denote $\mathfrak{s}_0$. With this, there is a natural identification of $H^2(M)$ with $\mathrm{Spin}^c(M)$. However we face the complication that in assuming that M bounds a rational homology 4-ball X, we cannot assume that X has a spin structure. The following result permits us to adapt to this possibility. (In addition to playing a role in considering splittings of classes in $\Theta^3_{\mathbb{Q}}$, in Section 13 we will use this result to extend a theorem from [Hedden et al. 2012] in which an added hypothesis was needed to ensure the existence of a spin structure on X.)

Theorem 6. *Suppose that $N_1 \# N_2 = \partial X$ for some smooth rational homology 4-ball X and that the order of $H_1(N_1)$ is odd. Then the image of the restriction map $\mathrm{Spin}^c(X) \to \mathrm{Spin}^c(N_1)$ contains the spin structure $\mathfrak{s}_0 \in \mathrm{Spin}^c(N_1)$. In particular, every element in the image of this restriction map is of the form $\alpha \cdot \mathfrak{s}_0$ for $\alpha \in \mathrm{Image}(H^2(X) \to H^2(N_1))$.*

Proof. Let $H = \mathrm{Image}(H^2(X) \to H^2(N_1))$, $S = \mathrm{Image}(\mathrm{Spin}^c(X) \to \mathrm{Spin}^c(N_1))$. As usual, the choice of an element $\mathfrak{s} \in S$ determines a bijection between H and S. In particular, the number of elements in S is the same as in H, which is odd. Conjugation defines an involution on S which commutes with restriction. Thus since S is odd, conjugation has a fixed point in S. But the only fixed element under conjugation is the spin structure, since $c_1(\bar{\mathfrak{s}}) = -c_1(\mathfrak{s})$. □

5. Basic obstructions from d-invariants

To each rational homology 3-sphere M and $\mathfrak{s} \in \mathrm{Spin}^c(M)$ there is associated an invariant $d(M, \mathfrak{s}) \in \mathbb{Q}$, defined in [Ozsváth and Szabó 2003]. It is additive under connected sum: $d(M \# N, (\mathfrak{s}_1, \mathfrak{s}_2)) = d(M\ , \mathfrak{s}_1) + d(N, \mathfrak{s}_2)$. A key result relating the d-invariant and bordism is the following, taken from the same reference.

Theorem 7. *If $M = \partial X$ with $H_*(X, \mathbb{Q}) \cong H_*(B^4, \mathbb{Q})$, and $\mathfrak{t} \in \mathrm{Spin}^c(X)$, then $d(M, \mathfrak{t}|_M) = 0$.*

5.1. *Obstruction theorem.* Suppose that $|H_1(M)|$ is odd and $\mathfrak{s}_0$ is the unique spin structure on M. For $\alpha \in H^2(M)$, we abbreviate $d(M, \alpha \cdot \mathfrak{s}_0)$ by $d(M, \alpha)$.

Definition 8. $\bar{d}(M, \alpha) = d(M, \alpha) - d(M, 0)$.

The following result will be sufficient to prove that $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$ is infinite.

Theorem 9. *Suppose $\{M_i\}$ is a set of 3-manifolds for which $H_1(M_i) = \mathbb{Z}_{m_i} \oplus \mathbb{Z}_{n_i}$, where the m_i and n_i are square-free and odd, and the elements of the full set $\{m_i, n_i\}$ are pairwise relatively prime. If a finite connected sum $\#_{k=1}^N \pm M_{i_k}$ represents a class in $\Theta^3_{\mathbb{Q}}$ that is in the image $\Phi(\bigoplus \Theta_p)$, then for all $i = i_k$, $1 \le k \le N$, and for all $(a, b) \in \mathbb{Z}_{m_i} \oplus \mathbb{Z}_{n_i}$,*

$$\bar{d}(M_i, (a, b)) = \bar{d}(M_i, (a, 0)) + \bar{d}(M_i, (0, b)).$$

Proof. Let $Y = \#_k \pm M_{i_k}$. We consider $k = 1$, abbreviating $M_{i_1} = M$ and $H_1(M) \cong \mathbb{Z}_m \oplus \mathbb{Z}_n$. Suppose that Y is in the image of Φ. Then $Y \# \bigoplus Y_{p_i} = \partial X$ for some collection of $\mathbb{Z}[p_i^{-1}]$-homology spheres Y_{p_i} and a rational homology ball X. Collecting summands, we can write $M \# N_m \# N_n \# N = \partial X$, where the prime factors of $|H_1(N_m)|$ all divide m, the prime factors of $|H_1(N_n)|$ all divide n, and $|H_1(N)|$ is relatively prime to mn. Let $(\mathfrak{s}_0, \mathfrak{s}_1, \mathfrak{s}_2, \mathfrak{s}_*) \in \text{Image}(\text{Spin}^c(X))$. (By Theorem 6 we can assume that the structure $\mathfrak{s}_0 \in \text{Spin}^c(M)$ is the spin structure.) Then, by Corollary 3, for all $a \in \mathbb{Z}_m$ and $b \in \mathbb{Z}_n$, there are elements $a' \in H_1(N_m)$ and $b' \in H_1(N_n)$ such that:

- $((a, 0) \cdot \mathfrak{s}_0, a' \cdot \mathfrak{s}_1, \mathfrak{s}_2, \mathfrak{s}_*) \in \text{Image}(\text{Spin}^c(X))$.
- $((0, b) \cdot \mathfrak{s}_0, \mathfrak{s}_1, b' \cdot \mathfrak{s}_2, \mathfrak{s}_*) \in \text{Image}(\text{Spin}^c(X))$.
- $((a, b) \cdot \mathfrak{s}_0, a' \cdot \mathfrak{s}_1, b' \cdot \mathfrak{s}_2, \mathfrak{s}_*) \in \text{Image}(\text{Spin}^c(X))$.

Thus we have the following vanishing conditions on the d-invariants:

- $d(M, \mathfrak{s}_0) + d(N_m, \mathfrak{s}_1) + d(N_n, \mathfrak{s}_2) + d(N, \mathfrak{s}_*) = 0$.
- $d(M, (a, 0) \cdot \mathfrak{s}_0) + d(N_m, a' \cdot \mathfrak{s}_1) + d(N_n, \mathfrak{s}_2) + d(N, \mathfrak{s}_*) = 0$.
- $d(M, (0, b) \cdot \mathfrak{s}_0) + d(N_m, \mathfrak{s}_1) + d(N_n, b' \cdot \mathfrak{s}_2) + d(N, \mathfrak{s}_*) = 0$.
- $d(M, (a, b) \cdot \mathfrak{s}_0) + d(N_m, a' \cdot \mathfrak{s}_1) + d(N_n, b' \cdot \mathfrak{s}_2) + d(N, \mathfrak{s}_*) = 0$.

Subtracting the second and third equality from the sum of the first and fourth yields

$$d(M, (a, b) \cdot \mathfrak{s}_0) - d(M, (a, 0) \cdot \mathfrak{s}_0) - d(M, (0, b) \cdot \mathfrak{s}_0) + d(M, \mathfrak{s}_0) = 0.$$

Recalling that $\bar{d}(M, \alpha)$ denotes $d(M, \alpha \cdot \mathfrak{s}_0) - d(M, \mathfrak{s}_0)$, this can be rewritten as

$$\bar{d}(M, (a, b)) - \bar{d}(M, (a, 0)) - \bar{d}(M, (0, b)) = 0.$$

Repeating for each M_i completes the proof of the theorem. $\square$

6. Lens space examples: $L(pq, 1)$.

Let (p_i, q_i) be pairs of relatively prime square-free odd integers such the products $p_i q_i$ are pairwise relatively prime. We prove:

Theorem 10. *No finite linear combination $\#_k \pm L(p_{i_k} q_{i_k}, 1)$ represents an element in the image $\Phi(\bigoplus \Theta_p) \subset \Theta^3_{\mathbb{Q}}$.*

Proof. To simplify notation, we use L_n to denote $L(n, 1)$. Assume there is such a finite linear combination. We consider the first term $L_{p_1 q_1}$ and simplify notation by writing $p = p_1$ and $q = q_1$. By Theorem 9 we would have for all $(a, b) \in \mathbb{Z}_p \oplus \mathbb{Z}_q$,

$$\bar{d}(L_{pq}, (a, b)) = \bar{d}(L_{pq}, (a, 0)) + \bar{d}(L_{pq}, (0, b)).$$

According to [Ozsváth and Szabó 2003], for some enumeration of Spin^c structures on $L(m, n)$, denoted $\mathfrak{s}_i$, $0 \le i < m$, if we let $D(m, n, i) = d(-L(m, n), \mathfrak{s}_i)$, there is the recursive formula

$$D(m, n, i) = \frac{mn - (2i + 1 - m - n)^2}{4mn} - D(n, m', i'),$$

where the primes denote reductions modulo n, $0 < n < m$, and $0 \le i < m$. The base case in the recursion is by definition $D(1, 0, 0) = 0$. For every Spin^c structure $\mathfrak{s}$ there is a conjugate structure $\bar{\mathfrak{s}}$ for which $d(M, \mathfrak{s}) = d(M, \bar{\mathfrak{s}})$ and $\mathfrak{s} \ne \bar{\mathfrak{s}}$ unless $\mathfrak{s}$ is the spin structure. We claim that for L_{pq}, the Spin^c structure $\mathfrak{s}_0$ does correspond to the spin structure. Indeed, an algebraic computation shows that $4pqD(pq, 1, i) = -4i^2 + 4pqi + pq(1 - pq)$, and in particular, $pqD(pq, 1, 0) = pq(1 - pq)$. The difference,

$$4pqD(pq, 1, i) - 4pqD(pq, 1, 0) = 4i(pq - i),$$

does not take on the value 0 for any $0 < i < pq$. Since the value of $D(pq, 1, 0)$ is unique among the d-invariants, it must correspond to the spin structure. In applying Theorem 9, we identify $\mathbb{Z}_p \oplus \mathbb{Z}_q \cong \mathbb{Z}_{pq}$, so that the pair $(a, b) \in \mathbb{Z}_p \oplus \mathbb{Z}_q$ corresponds to $aq + bp \in \mathbb{Z}_{pq}$. In this case, the criterion becomes

$$D(pq, 1, ap + bq) - D(pq, 1, ap) - D(pq, 1, bq) + D(pq, 1, 0) = 0.$$

Certainly $p + q < pq$, so we can apply the formula for D with $a = b = 1$. However, in this case, the sum is immediately calculated to equal -2, which is not 0. □

7. Infinite order examples

The examples of the previous section are sufficient to demonstrate that the quotient $\Theta^3_{\mathbb{Q}} / \Phi(\bigoplus \Theta_p)$ is infinite. We now present an argument to show it contains an infinitely generated free subgroup. To carry out this argument we need to make the additional assumption of primeness for the relevant p and q. Let $\{p_i, q_i\}$ be a set of distinct odd prime pairs with all elements distinct. We continue to denote $L(n,1)$ by L_n. This section is devoted to the proof of the following theorem.

Theorem 11. *The lens spaces $L_{p_i q_i}$ are linearly independent in the quotient group $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$.*

7.1. ***Notation.*** Suppose that $\sum_i b_i L_{p_i q_i} \subset \text{Image}\,(\Phi)$. We can assume that $b_1 \neq 0$. We simplify notation, writing p and q for p_1 and q_1, respectively. There is no loss of generality in assuming that for all i, $b_i = 4k_i$ for some k_i, and write $k = k_1$.

Following our earlier approach, we will show that a contradiction arises from the assumption that $N = 4kL_{pq} \# M_p \# M_q \# M_* = \partial X$ for some rational homology 4-ball X, where the orders of $H_1(M_p)$ and $H_1(M_q)$ are powers of p and q, respectively, and the order of $H_1(M_*)$ is relatively prime to pq.

According to Theorem 4, the p-primary part of the associated metabolizer, $\mathcal{M}_p$, includes a vector $A = ((1,\dots,1,\alpha_{2k+1},\dots,\alpha_{4k}),g) \in (\mathbb{Z}_p)^{4k} \oplus H_1(M_p)$. Similarly the q-primary part of the associated metabolizer, $\mathcal{M}_q$, includes a vector $B = ((1,\dots,1,\beta_{2k+1},\dots,\beta_{4k}),h) \in (\mathbb{Z}_q)^{4k} \oplus H_1(M_q)$.

7.2. ***Constraints on the d-invariants.*** We let the spin structures on $L(pq,1)$, M_p, and M_q be $\mathfrak{s}_0, \mathfrak{s}_0'$ and $\mathfrak{s}_0''$, respectively. Consider now the vectors 0, aA, bB, and $aA + bB \in \mathcal{M}$. Computing the d-invariant associated to each, we find that each of the following sums is 0:

$$2kd(L_{pq},s_0) + \sum_{i=2k+1}^{4k} d(L_{pq},\mathfrak{s}_0) + d(M_p,\mathfrak{s}_0') + d(M_q,\mathfrak{s}_0'') + d(M_*,\mathfrak{t}),$$

$$2kd(L_{pq},aq\cdot s_0) + \sum_{i=2k+1}^{4k} d(L_{pq},aq\alpha_i\cdot\mathfrak{s}_0) + d(M_p,ag\cdot\mathfrak{s}_0') + d(M_q,\mathfrak{s}_0'') + d(M_*,\mathfrak{t}),$$

$$2kd(L_{pq},bp\cdot s_0) + \sum_{i=2k+1}^{4k} d(L_{pq},bp\beta_i\cdot\mathfrak{s}_0) + d(M_p,\mathfrak{s}_0') + d(M_q,bh\cdot\mathfrak{s}_0'') + d(M_*,\mathfrak{t}),$$

$$2kd(L_{pq},(aq+bp)\cdot s_0) + \sum_{i=2k+1}^{4k} d(L_{pq},(aq\alpha_i+bp\beta_i)\cdot\mathfrak{s}_0) + d(M_p,ag\cdot\mathfrak{s}_0') + d(M_q,bh\cdot\mathfrak{s}_0'') + d(M_*,\mathfrak{t}).$$

Note. We have again used that the inclusion $\mathbb{Z}_p \subset \mathbb{Z}_{pq}$ takes α to αq, and similarly for $\mathbb{Z}_q$ and β. We now take the sum of the first and last equation, and subtract the sum of the middle two. As before, we continue abbreviating $d(M,a\cdot s_0)$ by $d(M,a)$. The result is that, for some set of a_i and b_i,

$$2k\big(d(L_{pq},aq+bp) - d(L_{pq},aq) - d(L_{pq},bp) + d(L_{pq},0)\big) + \sum_{i=2k+1}^{4k}\big(d(L_{pq},a_iq+b_ip) - d(L_{pq},a_iq) - d(L_{pq},b_ip) + d(L_{pq},0)\big) = 0.$$

We now introduce further notation. Let

$$\delta(L_{pq}, a, b) = d(L_{pq}, aq + bp) - d(L_{pq}, aq) - d(L_{pq}, bp) + d(L_{pq}, 0).$$

With this, we have proved the following lemma.

Lemma 12. *If the lens spaces $L_{p_i q_i}$ are linearly dependent in $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta^3_p)$, and, for $p = p_1$ and $q = q_1$, L_{pq} has nonzero coefficient in some linear relation, then for all a and b there are k, a_i and b_i such that*

$$2k\delta(L_{pq}, a, b) + \sum_{i=2k+1}^{4k} \delta(L_{pq}, a_i, b_i) = 0.$$

7.3. ***Computation of bounds on $\delta(L_{pq}, a, b)$.*** Note that $\delta(L_{pq}, a, b) = 0$ if $a = 0$ or $b = 0$. Given Lemma 12, the proof of Theorem 11 is completed with the following result.

Lemma 13. *For all $a \neq 0 \mod p$ and $b \neq 0 \mod q$, $\delta(L_{pq}, a, b) < 0$.*

Proof. All Spin^c structures are included by considering the range

$$-(p-1)/2 \le a \le (p-1)/2 \quad \text{and} \quad -(q-1)/2 \le b \le (q-1)/2.$$

By symmetry we can exclude the case $a < 0$. Since the formula for the d-invariant $d(L_{pq}, i)$ assumes $i \ge 0$, there are three cases to consider.

(1) $a > 0$, $b > 0$.

(2) $a > 0$, $-aq/p < b < 0$.

(3) $a > 0$, $b < -aq/p$.

The formula for the d-invariant in the current case is

$$4n(d(L_n, i)) = n - (2i + 1 - n - 1)^2 = n - n^2 + 4ni - 4i^2$$

for $0 \le i < n$. We now compute $4pq\delta(L_{pq}, aq + bp)$ in each of the three cases. First note that

$$\delta(L_{pq}, aq + bp) = d(L_{pq}, aq + bp) - d(L_{pq}, aq) - d(L_{pq}, bp) + d(L_{pq}, 0).$$

Applying the formula for the d-invariant, taking care that i is positive in the calculation of $d(L_n, i)$ we find in the first case, $\delta = -8abpq$, which is negative. In the second case we compute $\delta = -8b(a - p)pq$, which is again negative (since $b < 0$ and $a < (p-1)/2$). In the third case, $\delta = -8apq(b + q)$, which is negative since $b > -(q-1)/2$. This completes the proof. □

a	$b=0$	1	2	3	4	5	6
2	–52	–112	–32	–72	28	8	–2
1	52	–8	–58	32	–128	18	–28
0	0	70	20	–20	80	–70	–80
–1	52	–8	72	32	2	112	–28
–2	–52	18	–32	58	28	8	128

Table 1. $65\,d(L(65,8),13a+5b)$.

a	$b=0$	1	2	3	4	5	6
2	0	–2	0	0	0	2	2
1	0	–2	–2	0	–4	0	0
0	0	0	0	0	0	0	0
–1	0	–2	0	0	–2	2	0
–2	0	0	0	2	0	2	4

Table 2. $d(L(65,8),13a+5b)-d(L(65,8),13a)-d(L(65,8),5b)$.

8. An order-2 lens space that does not split

We now consider a lens space that represents 2-torsion in $\Theta^3_{\mathbb{Q}}$. Let $M = L(65,8)$; since $8^2 = -1 \mod 65$, we have $M = -M$ and $2M = 0 \in \Theta^3_{\mathbb{Q}}$. We show that M does not split. It follows quickly from the fact that $L(65,8)$ is of finite order in $\Theta^3_{\mathbb{Q}}$ that, for the spin structure $\mathfrak{s}^*$, $d(L(65,8),\mathfrak{s}^*)=0$. One can compute directly from the formula for D given above that the value 0 is realized only by $\mathfrak{s}_{36}$. Thus, in applying Theorem 9, we identify the homology class $x \in H_1(L(65,8))$ with the Spin^c structure $\mathfrak{s}_{36+x}$, where the index is taken modulo 65. Table 1 presents the values of $d(L(65,8),13a+5b)$ (multiplied by 65 to clear denominators). Rows correspond to the values of a and columns to b. The central row and left column correspond to $a=0$ and $b=0$, respectively. Symmetry permits us to list only the values with $b \geq 0$. In Table 2 we list the differences, $d(L(65,8),13a+5b) - d(L(65,8),13a) - d(L(65,8),5b)$, with the nonzero entries demonstrating the failure of additivity.

9. Infinite 2-torsion

We now generalize the previous example to describe an infinite subgroup of $\Theta^3_{\mathbb{Q}}$ consisting of 2-torsion that injects into the quotient $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$. Consider the family $N_n = L(4(5n+1)^2+1, 2(5n+1))$; for $n=-1$ we have $-L(65,8)$ as in the previous section, but we simplify the computations by restricting to $n>0$. Expanding, we have $N_n = L(5(20n^2+8n+1), 2(5n+1))$. If $n \neq 3 \mod 5$, then

$20n^2+8n+1$ is not divisible by 5. By Appendix A we can further assume that the number n are selected so that n is divisible by 5, and the integers $20n^2+8n+1$ are pairwise relatively prime and square-free. We enumerate the set of such n as n_i and abbreviate the corresponding lens spaces as $L(5p_i, q_i) = N_{n_i}$. The remainder of this section is devoted to proving the following.

Theorem 14. *The set $\{N_{n_i}\}$ generates an infinite subgroup consisting of elements of order 2 in $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus \Theta_p)$.*

To begin, we need to identify the spin structure. We use the recursion formula

$$D(m,n,i) = \frac{mn-(2i+1-m-n)^2}{4mn} - D(n,m',i')$$

to compute relevant d-invariants. We are interested in the lens spaces $L(4r^2+1, 2r)$. One step of the recursion reduces this to $L(2r, 1)$, and another step reduces it to S^3. Since we need to reduce modulo $2r$, for $0 \le i < 4r^2+1$, let y be the remainder of i modulo $2r$ and x the quotient so that $2rx + y = i$. So we write Spin^c structures as $\mathfrak{s}_{2rx+y}$ for $0 \le y < 2r$ and $0 \le 2rx + y < 4r^2+1$. Carrying out the arithmetic yields:

Lemma 15. *For any $r > 0$, x and y with $0 \le y < 2r$ and $0 \le 2rx + y < 4r^2+1$:*

(1) $d(L(4r^2+1, 2r), \mathfrak{s}_{2rx+y}) = \dfrac{2\big(rx^2+(y-r(2r+1))x-r(y^2-(2r-1)y-r)\big)}{4r^2+1}$.

(2) *The discriminant of the numerator, viewed as a quadratic polynomial in the variable x, is $4(y-r)^2(4r^2+1)$. Moreover it is the square of an integer if and only if $y = r$.*

(3) $d(L(4r^2+1, 2r), \mathfrak{s}_{2rx+y}) = 0$ *if and only if $x = r$ and $y = r$.*

(4) *The spin structure on $L(4r^2+1, 2r)$ is $\mathfrak{s}_{2r^2+r}$.*

In our case $r = 5n+1$ and the spin structure is $\mathfrak{s}_{50n^2+25n+3}$.

Proof of Theorem 14. For each n, we write $N_n = L(5p_n, q_n)$ and assume that some linear combination $\sum N_{n_i} = 0 \in \Theta^3_{\mathbb{Q}}/\Phi(\bigoplus_{p\in\mathcal{P}} \Theta^3_{\mathbb{Z}[1/p]})$. We write the first term in the sum as $N = L(5p, q)$, where $p = 20n^2+8n+1$. Since the sum splits, for some collection of primes r_j and manifolds M_{r_j} with $H_1(M_{r_j})$ r_j-torsion, we have

$$N \mathbin{\#} \mathop{\#}_{i>1} N_{n_i} \mathbin{\#} \mathop{\#}_{j} M_{r_j} = \partial X,$$

where X is a rational homology ball. We can collect terms as $N \# M_p \# M_m = \partial X$ where M_p includes all the M_{r_j} for which r_j divides p, and M_m contains all the other summands, including all the N_{n_i} with $i > 1$.

The homology of this connected sum of three-manifolds splits into the direct sum of three groups: $(\mathbb{Z}_5 \oplus \mathbb{Z}_p) \oplus G_p \oplus G_m$, where the order of G_p is a product of prime factors of p, 5 does not divide the order of G_p, and the orders of G_p and G_m are relatively prime. It follows that the 5-torsion in the metabolizer, $\mathcal{M}_5$, is contained in $(\mathbb{Z}_5, 0) \oplus 0 \oplus G_m$. The direct sum of all primary parts of the metabolizer for primes that divide p, $\mathcal{M}_p$, is contained in $\mathcal{M}_p = (0, \mathbb{Z}_p) \oplus G_p \oplus 0$.

As in our previous arguments, $\mathcal{M}_5$ contains an element of the form $(1, 0) \oplus 0 \oplus a''$, and $\mathcal{M}_p$ contains an element $(0, 1) \oplus b'' \oplus 0$. Continuing as in the early proofs, we find that for all a and b,

$$\bar{d}(L(5p, q), (a, b)) = \bar{d}(L(5p, q), (a, 0)) + \bar{d}(L(5p, q), (0, b)).$$

Or, writing $\mathbb{Z}_5 \oplus \mathbb{Z}_p$ as $\mathbb{Z}_{5p}$,

$$\bar{d}(L(5p, q), pa + 5b) = \bar{d}(L(5p, q), pa) + \bar{d}(L(5p, q), 5b).$$

Since $L(5p, q)$ is of order two, for the spin structure, the d-invariant vanishes so the $\bar{d}$-invariant is the same as the d-invariant. We let $a = 1$ and $b = -1$ and arrive at a contradiction by showing the following equality does not hold:

$$d(L(5p, q), p - 5) = d(L(5p, q), p) + d(L(5p, q), -5).$$

To apply Lemma 15 we need to express each of

$$(50n^2 + 25n + 3) + p - 5, \quad (50n^2 + 25n + 3) + p \quad \text{and} \quad (50n^2 + 25n + 3) - 5$$

as $2(5n + 1)x + y$. Simple algebra yields the following pairs (x, y) for these three respective Spin^c structures:

- $a = 1, b = -1$: $\quad (x, y) = (7n + 1, 9n - 3)$.
- $a = 1, b = 0$: $\quad (x, y) = (7n + 1, 9n + 2)$.
- $a = 0, b = -1$: $\quad (x, y) = (5n + 1, 5n - 4)$.

Finally one uses these expressions to determine that for all n

$$d(L(5p, q), p - 5) - d(L(5p, q), p) - d(L(5p, q), -5) = 4.$$

Since the difference is not zero, no splitting exists and the proof of Theorem 14 is complete. □

10. Topologically split examples

In this section, we apply Theorem 9 to find examples of manifolds that split topologically but not smoothly. We begin by carefully examining an example in which the splitting exists smoothly, focusing on the computation of the d-invariants and next illustrate the modifications which do not change its topological cobordism class but alter it smoothly. The deepest aspect of the work is in the determination of the d-invariants. In brief, the manifold we look at is 15-surgery on the $(3,5)$-torus knot, $T_{3,5}$, denoted by $S^3_{15}(T_{3,5})$. This is homeomorphic to the connected sum $L(3,5) \# -L(5,3)$. Next, letting D denote the untwisted double of the trefoil knot ($D = \mathrm{Wh}(T_{2,3})$), which is topologically slice, we consider $S^3_{15}(T_{3,5} \# D)$ and prove that it does not split in the cobordism group.

In this section and the next, and also in Appendix B, we develop properties of the Heegaard–Floer complex of specific torus knots and tensor products of certain of these complexes. More extensive computations appear in [Hancock et al. 2013].

10.1. $\bar{d}(S^3_{15}(T_{3,5}), i)$. We now determine the doubly filtered Heegaard–Floer complex $\mathrm{CFK}^\infty(S^3, T_{3,5})$. This complex is by definition a doubly filtered, graded chain complex over $\mathbb{F}_2$. Thus a set of filtered generators can be illustrated on a grid with the coordinates representing the filtration levels and the grading marked. There is an action of $\mathbb{Z}$ on the complex, and if we let U be the generator, this makes the complex a $\mathbb{F}_2[U, U^{-1}]$-module. The action of U on the complex lowers filtration levels by 1 and gradings by 2.

We now show that $\mathrm{CFK}^\infty(S^3, T_{3,5})$ is as illustrated in Figure 2. In order to find this decomposition, we start by focusing on the central column (for which the top-most generator is at filtration level $j = 4$ and is labeled with its grading 0). The vertical column, $i = 0$, represents the subquotient complex $\widehat{\mathrm{CFK}}(S^3, T_{3,5})$. We begin by explaining why it appears as it does in the illustration. According to [Ozsváth and Szabó 2005, Theorem 1.2], since for torus knots there is an integer surgery that yields a lens space, $\widehat{\mathrm{HFK}}(S^3, T_{3,5}, j)$, the quotient of the j-filtration level by the $(j-1)$-filtration level is completely determined by the Alexander polynomial,

$$\Delta_{T_{3,5}}(t) = 1 - (t^{-1} + t) + (t^{-3} + t^3) - (t^{-4} + t^4).$$

This explains the location of the generators of $\widehat{\mathrm{CFK}}(S^3, T_{3,5})$. Similarly the same work determines the grading of the generators. The fact the complex $\widehat{\mathrm{CFK}}(S^3, T_{3,5})$ is a filtration of the complex $\widehat{CF}(S^3)$, which has homology $\mathbb{F}_2$ with its generator at grading level 0, forces the vertical arrows, presenting the boundary maps, to be as illustrated. To build the CFK^∞ diagram from the $\widehat{\mathrm{CFK}}$ diagram, we first apply the action of U to fill in the generators as well as the all the vertical arrows. We next note that the homology groups $\widehat{\mathrm{HFK}}(T_{3,5}, i)$ can be computed using the horizontal slice $j = 0$ instead of the vertical slice, and this forces the existence of the horizontal arrows as drawn. With this much of the diagram drawn and the action of U lowering grading by 2, the gradings of all the elements in the diagram are determined. Finally we note that the fact that the boundary map lowers gradings by 1 rules out the possibility of any other arrows.

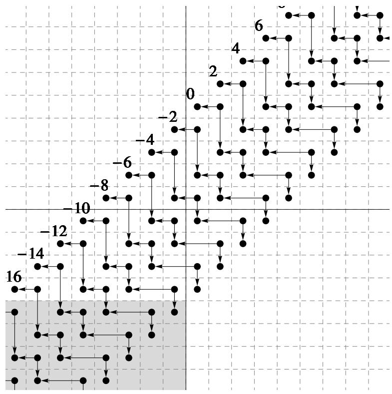

Figure 2. The case $s = -4$ with the quotienting subgroup shaded.

According to [Ozsváth and Szabó 2004], the complex $\mathrm{CFK}^+(S^3_{15}(T_{3,5}), s)$, for $-7 \le s \le 7$, is given by the quotient

$$\mathrm{CFK}^\infty(S^3, T_{3,5}) / \mathrm{CFK}^\infty(S^3, T_{3,5})_{i<0, j<s}[-\eta],$$

where η is a grading shift:

$$\eta = \frac{-(2s-15)^2 + 15}{60}.$$

Figure 2 illustrates the case $s = -4$ with the quotienting subgroup shaded in the diagram. By definition, the d-invariant is the minimal grading among all classes in the group $\mathrm{HFK}^+(S^3_{15}(T_{3,5}), s)$, which are in the image of U^n for all n. From the diagram, without shifting the gradings, we see this minimum for $\mathrm{HFK}^+(S^3_{15}(T_{3,5}), -4)$ is -8: one generator of grading level -10 has been killed, and all such generators are homologous. The values for all Spin^c structures, $s = -7, -6, \ldots, 6, 7$ are given in order as

$$\{-14, -12, -10, -8, -8, -6, -4, -4, -2, -2, -2, 0, 0, 0, 0\}.$$

After the grading shift, the values are all of the form $a_i/30$, where, in order, the a_i are

$$\{-7, -3, 5, 17, -27, -7, 17, -15, 17, -7, -27, 17, 5, -3, -7\}.$$

Finally to compute $\bar{d}$, we subtract $-15/30$ (the value for the spin structure) from each entry and find that the values of $\bar{d}$ are given by $b_i/30$ for the following values of b_i in order:

$$\{8, 12, 20, 32, -12, 8, 32, 0, 32, 8, -12, 32, 20, 12, 8\}.$$

We have listed these values in Table 3, in which we write each value of s as $5a + 3b$ mod 15 for $-1 \le a \le 1$ and $-2 \le b \le 2$.

a	$b = -2$	-1	0	1	2
1	32	8	**20**	8	32
0	**12**	**–12**	**0**	**–12**	**12**
–1	32	8	**20**	8	32

Table 3. $30\,\bar{d}(S^3_{15}(T_{3,5}), 5a + 3b)$.

Since $S^3_{15}(T_{3,5})$ is the connected sum of lens spaces, Theorem 9 predicts a pattern in the chart: each element should be the sum of the entries of its projection on the main axes. This is the case. Notice for instance that the top-right entry 32 in position $(a, b) = (1, 2) \in \mathbb{Z}_3 \oplus \mathbb{Z}_5$ (which represents $1(5) + 2(3) = 11 \in \mathbb{Z}_{15}$) is the sum of the entries in positions $(2, 0)$ and $(0, 1)$, 12 and 20, respectively.

10.2. $\bar{d}(S^3_{15}(T_{3,5}\#D), \mathfrak{s})$. In order to compute the $\bar{d}$-invariants that are associated to surgery on the connected sum, we first must compute CFK^∞ for the connected sum of knots. The complex $\mathrm{CFK}^\infty(T_{2,3})$ is illustrated in Figure 3, left, and it follows from [Hedden 2007] that, modulo acyclic subcomplexes, the chain complex for D is the same. (The results of the same article are focused on the hat complex $\widehat{\mathrm{CFK}}(K)$ but extend to the full CFK^∞ complex. For more details, see [Hedden et al. 2014].)

At this point we need to analyze the tensor product

$$C = \mathrm{CFK}^\infty(T_{3,5}) \otimes_{\mathbb{F}[U,U^{-1}]} \mathrm{CFK}^\infty(T_{2,3}).$$

This complex is fairly complicated, containing 21 generators, but it is easily seen that it contains a subcomplex C' as illustrated in Figure 3, right. This subcomplex

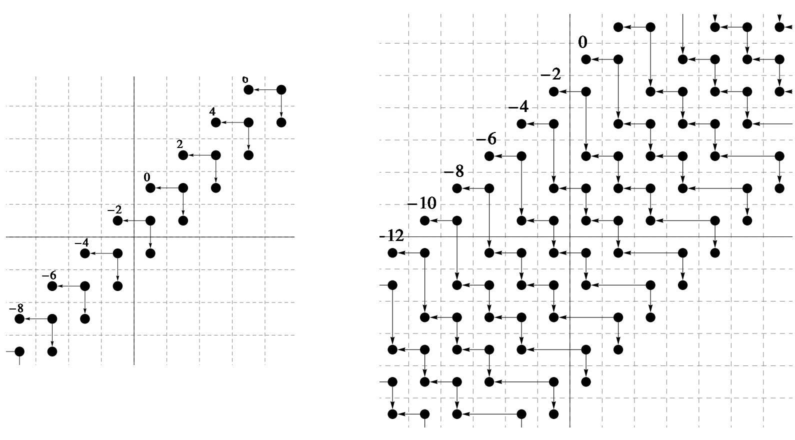

Figure 3. Left: the complex $\mathrm{CFK}^\infty(T_{2,3})$. Right: The subcomplex C' of $\mathrm{CFK}^\infty(T_{3,5}) \otimes_{\mathbb{F}[U,U^{-1}]} \mathrm{CFK}^\infty(T_{2,3})$.

a	$b=-2$	-1	0	1	2
1	$\underline{-28}$	8	**20**	8	$\underline{-28}$
0	**12**	**–12**	**0**	**–12**	**12**
–1	$\underline{-28}$	8	**20**	8	$\underline{-28}$

Table 4. $30\,\bar{d}(S^3_{15}(T_{3,5}\,\#\,D), 5a+3b)$.

carries the homology of the overall complex but does not contain all generators of a given grading. However, by examining the full complex with 21 generators, we have the following observation.

The complex $C_{i<m,j<n}$ contains a generator of grading 0 if and only if $C'_{i<m,j<n}$ contains a generator of grading 0. In particular, d-invariants for C can be computed using C'.

Using this diagram to compute the minimal gradings of classes in

$$\mathrm{CFK}^\infty(T_{3,5}\,\#\,D)/\,\mathrm{CFK}^\infty(T_{3,5}\,\#\,D)_{i<0,j<s}$$

for $-7 \le s \le 7$, we get the following:

$$\{-14, -12, -10, -10, -8, -6, -6, -4, -4, -2, -2, -2, 0, 0, 0\}.$$

After shifting gradings by $-\eta$, the values are of the form $a_i/30$, where the a_i are, in order,

$$\{-7, -3, 5, -43, -27, -7, -43, -15, -43, -7, -27, -43, 5, -3, -7\}.$$

To compute $\bar{d}$, we add $15/30$ to each term, yielding the values $b_i/30$, where the b_i are

$$\{8, 12, 20, -28, -12, 8, -28, 0, -28, 8, -12, -28, 20, 12, 8\}.$$

We can arrange these in a chart as in Table 4.

Notice that the entries on the axes are unchanged, but the underlined entries are no longer the sum of the values of the projections; that is, $-28 \neq 12 + 20$. Thus, according to Theorem 9, this manifold is not $\mathbb{Q}$-homology cobordant to any manifold of the form $M_3\,\#\,M_5\,\#\,M_q$.

10.3. ***Second example.*** As a second example we consider the case of $S^3_{35}(T_{5,7})$ and $S^3_{35}(T_{5,7}\,\#\,D)$ and illustrate the analogous charts as above (this time multiplied by 70 to clear denominators). The first chart, Table 5, necessarily demonstrates additivity; the second, in Table 6, upon examination does not. This becomes more apparent by considering the third chart, in Table 7, formed as the difference of the first two, but not multiplied by 70. The underlined entries illustrate the failure of additivity. Considering this difference is a simplifying approach of the general proof in the next section.

a	$b = -3$	-2	-1	0	1	2	3
2	–68	–108	–48	**–28**	–48	–108	–68
1	–12	–52	8	**28**	8	–52	–12
0	**–40**	**–80**	**–20**	**0**	**–20**	**–80**	**–40**
–1	–12	–52	8	**28**	8	–52	–12
–2	–68	–108	–48	**–28**	–48	–108	–68

Table 5. $70\,\bar{d}(S^3_{35}(T_{5,7}), 7a+5b)$.

a	$b = -3$	-2	-1	0	1	2	3
2	$\underline{72}$	$\underline{32}$	92	**112**	92	$\underline{32}$	$\underline{72}$
1	128	88	8	**28**	8	88	128
0	**100**	**60**	**–20**	**0**	**–20**	**60**	**100**
–1	128	88	8	**28**	8	88	128
–2	$\underline{72}$	$\underline{32}$	92	**112**	92	$\underline{32}$	$\underline{72}$

Table 6. $70\,\bar{d}(S^3_{35}(T_{5,7}\,\#\,D), 7a+5b)$.

a	$b = -3$	-2	-1	0	1	2	3
2	$\underline{2}$	$\underline{2}$	2	2	2	$\underline{2}$	$\underline{2}$
1	2	2	0	0	0	2	2
0	2	2	0	0	0	2	2
–1	2	2	0	0	0	2	2
–2	$\underline{2}$	$\underline{2}$	2	2	2	$\underline{2}$	$\underline{2}$

Table 7. $\bar{d}(S^3_{35}(T_{5,7}\,\#\,D), 7a+5b) - \bar{d}(S^3_{35}(T_{5,7}), 7a+5b)$.

11. Topologically split examples, general case

We now wish to generalize the examples of the previous section. To do so, we begin by choosing an infinite set of integers $\{p_i\}$ with the following properties: (1) all the p_i are odd; (2) the elements of the full set $\{p_i\} \cup \{p_i + 2\}$ are pairwise relatively prime; and (3) each p_i and $p_i + 2$ is square-free. The existence of such a set is demonstrated in Appendix A, and throughout this section we assume all p are selected from this set. In the previous example we needed to track grading shifts. It will simplify our discussion if we avoid dealing the grading shifts as follows. Define $\tilde{d}(S^3_n(K), s) = d(S^3_n(K), s) + \eta$. That is, $\tilde{d}$ is computed as is the d-invariant, except without the grading shift, the induced grading on

$$\mathrm{CFK}^+(S^3_N(K), s) = \mathrm{CFK}^\infty(S^3, K)/\,\mathrm{CFK}^\infty(S^3, K)_{\{i<0,\,j<s\}}.$$

Since p is odd, we can write $p = 2n+1$ and let $q = p+2 = 2n+3$. Our manifolds of interest are $S^3_{pq}(T_{p,q})$ and $S^3_{pq}(T_{p,q} \# D)$. We collect here the results of a few elementary calculations.

Theorem 16.

(1) *The surgery coefficient is* $pq = 4n^2 + 8n + 3$.

(2) *The three-genus satisfies*

$$g(T_{p,q}) = 2n(n+1) = 2n^2 + 2n \quad \textit{and} \quad g(T_{p,q} \# D) = 2n^2 + 2n + 1.$$

(3) Spin^c *structures are parameterized by* s, *with*

$$-(2n^2 + 4n + 1) \le s \le (2n^2 + 4n + 1).$$

(4) *Generators of* $\widehat{\mathrm{CFK}}(T_{p,q})$ *have filtration level* j, *where*

$$-2n(n+1) \le j \le 2n(n+1).$$

The main result of this section is the following.

Theorem 17. $\bar{d}(S^3_{pq}(T_{p,q} \# D), s)$ *does not satisfy additivity as given in Theorem 9.*

Proof. The space $S^3_{pq}(T_{p,q})$ satisfies the additive property as in Theorem 9. Suppose that $S^3_{pq}(T_{p,q} \# D)$ also satisfies the additivity property. Then the difference $\bar{d}(S^3_{pq}(T_{p,q}), (a,b)) - \bar{d}(S^3_{pq}(T_{p,q} \# D), (a,b))$ also satisfies the additivity property. We denote this difference by $\bar{d}'(a,b)$ or $\bar{d}'(aq+bp)$. Note that it is unnecessary to add the grading shift η to the amount we get from the diagram when computing either of the values $\bar{d}(S^3_{pq}(T_{p,q}), (a,b))$ or $\bar{d}(S^3_{pq}(T_{p,q} \# D), (a,b))$ since they have the same grading shift. Namely

$$\begin{aligned}\bar{d}'(a,b) = \tilde{d}(S^3_{pq}(T(p,q)), (a,b)) &- \tilde{d}(S^3_{pq}(T_{p,q} \# D), (a,b)) \\ - \tilde{d}(S^3_{pq}(T_{p,q}), 0) &+ \tilde{d}(S^3_{pq}(T_{p,q} \# D), 0).\end{aligned}$$

From our choice of p and q, we have $(n+1)p + (-n)q = 1$. Thus the additivity property implies the equality

$$\bar{d}'(1) = \bar{d}'((n+1)p) + \bar{d}'(-nq),$$

or, equivalently,

$$\begin{aligned}\text{(11-1)}\quad & \tilde{d}(S^3_{pq}(T_{p,q}), 1) - \tilde{d}(S^3_{pq}(T_{p,q} \# D), 1) \\ &= \tilde{d}(S^3_{pq}(T_{p,q}), (n+1)p) - \tilde{d}(S^3_{pq}(T_{p,q} \# D), (n+1)p) \\ &\quad + \tilde{d}(S^3_{pq}(T_{p,q}), -nq) - \tilde{d}(S^3_{pq}(T_{p,q} \# D), -nq) \\ &\quad - \tilde{d}(S^3_{pq}(T_{p,q}), 0) + \tilde{d}(S^3_{pq}(T_{p,q} \# D), 0).\end{aligned}$$

Since $(n+1)p = 2n^2+3n+1$ lies between the genus of $T(p,q)$ (and of $T_{p,q} \# D$) and the upper bound on the parameters for the Spin^c structures,

$$2n^2+2n+1 < 2n^2+3n+1 < 2n^2+4n+1,$$

the values of the $\tilde{d}$-invariants are easily seen to be 0. On the other hand, the number $-nq$ is greater than the lower bound on the parameters for the Spin^c structures and less than the negative of the genus,

$$-(2n^2+4n+1) < -(2n^2+3n) < -(2n^2+2n+1),$$

and thus one sees that the $\tilde{d}$-invariants associated to $-nq$ take the same value $-2s = 2(2n^2+3n)$ for both $T_{p,q}$ and $T_{p,q} \# D$. Thus, in contradicting additivity, it remains to show that the equality

$$\tilde{d}(S^3_{pq}(T_{p,q}),1) - \tilde{d}(S^3_{pq}(T_{p,q} \# D),1) = -\tilde{d}(S^3_{pq}(T_{p,q}),0) + \tilde{d}(S^3_{pq}(T_{p,q} \# D),0)$$

does not hold.

Now we will compute $\tilde{d}$ of both spaces for Spin^c structures 0 and 1. Observe that within width 1 from the diagonal $j=i$, the complex $\mathrm{CFK}^\infty(S^3,T_{p,q})$ looks like $\mathrm{CFK}^\infty(S^3,T_{2,3})$ if n is odd or $\mathrm{CFK}^\infty(S^3,T_{2,5})$ if n is even. This depends on the fact that near the origin the complex $\mathrm{CFK}^\infty(S^3,T_{p,q})$ looks like that of the $(2,k)$-torus knots. In Appendix B we prove that the Alexander polynomial of $T_{p,p+2}$ is of the form $1+\sum_{i>0} a_i(t^{-i}+t^i)$, where $a_i = \pm 1$ for $i \le (p-1)/2$. As in the example of the previous section, this determines the "zig-zag" feature of the CFK^∞ complex near the origin. Tensoring with the trefoil complex does not alter this pattern.

The generators of the same grading $2l$ of $[x,-1,0]$ if n is odd (or $[x,0,0]$ if n is even) lies above the antidiagonal $i+j=-1$ (or $i+j=0$). So, in order to compute $\tilde{d}(S^3_{pq}(T(p,q)),s)$ for $s=0,1$, we may assume in the computations that the complex we are considering is one of

$$\begin{cases} \mathrm{CFK}^\infty(S^3,T_{2,3}) & \text{if } n \text{ is odd,} \\ \mathrm{CFK}^\infty(S^3,T_{2,5}) & \text{if } n \text{ is even.} \end{cases}$$

It is now easy to compute

$$\tilde{d}(S^3_{pq}(T_{p,q}),s) = \begin{array}{c|cc} s & n \text{ odd} & n \text{ even} \\ \hline 1 & 2l+2 & 2l \\ 0 & 2l & 2l \end{array}$$

Near the diagonal $j=i$, the complex $\mathrm{CFK}^\infty(S^3,T_{p,q} \# D)$ looks like

$$\begin{cases} \mathrm{CFK}^\infty(S^3,T_{2,5}) & \text{if } n \text{ is odd,} \\ \mathrm{CFK}^\infty(S^3,T_{2,3})[-2] & \text{if } n \text{ is even.} \end{cases}$$

The grading of $[x,-1,0]$ is $2l-2$ if n is even and the grading of $[x,0,0]$ is $2l$ if n is odd. Thus we have

$$\tilde{d}(S^3_{pq}(T_{p,q}\,\#\,D),s) = \begin{array}{c|cc} s & n \text{ odd} & n \text{ even} \\ \hline 1 & 2l & 2l \\ 0 & 2l & 2l-2 \end{array}$$

We see that

$$\tilde{d}(S^3_{pq}(T_{p,q}),s) - \tilde{d}(S^3_{pq}(T_{p,q}\,\#\,D),s) = \begin{array}{c|cc} s & n \text{ odd} & n \text{ even} \\ \hline 1 & 2 & 0 \\ 0 & 0 & 2 \end{array}$$

This shows that (11-1) cannot be satisfied. We conclude that the space $S^3_{pq}(T_{p,q}\# D)$ does not satisfy the additive property of Theorem 9. $\square$

11.1. ***The image of*** $\mathcal{K}$ ***in*** $\boldsymbol{\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus\Theta_p)}$ ***is infinite.*** This is a consequence of the following result.

Theorem 18. *The spaces* $N_{p,q} = S^3_{pq}(T_{p,q}\,\#\,D)\,\#-S^3_{pq}(T_{p,q}) \in \mathcal{K}$ *are distinct in the quotient* $\Theta^3_{\mathbb{Q}}/\Phi(\bigoplus\Theta_p)$.

Proof. Observe that $S^3_{pq}(T_{p,q}\# D)\#-S^3_{pq}(T_{p,q}) \in \mathcal{K}$ since the knots are topologically concordant. We next observe that these manifolds have the property that no linear combination with all coefficients ± 1 is trivial in the quotient. Suppose that some such linear combination was trivial. Then focusing on any particular pair (p,q), we would have that $S^3_{pq}(T_{p,q}\,\#\,D)\,\#\,M_p\,\#\,M_q\,\#\,M_m = \partial X$ for a rational homology ball X, where the order of M_p is a product of prime factors of p, the order of M_q is a product of prime factors of q, and the order of M_m is relatively prime to pq. (This uses the fact that $S^3_{pq}(T_{p,q})$ does split as a connected sum.)

The existence of this connected sum decomposition implies the additivity for d-invariants of $S^3_{pq}(T_{p,q}\,\#\,D)$ in a way that contradicts Theorem 17. $\square$

12. Knot concordance

We denote by $\mathcal{C}$ the classical smooth knot concordance group. Levine [1969] defined the algebraic concordance group $\mathcal{G}$ and the rational algebraic concordance group, $\mathcal{G}^{\mathbb{Q}}$. He also defined a surjective homomorphism $\mathcal{C} \to \mathcal{G}$, proved that the natural map $\mathcal{G} \to \mathcal{G}^{\mathbb{Q}}$ is injective, and proved that $\mathcal{G}^{\mathbb{Q}}$ is isomorphic to an infinite direct sum of groups isomorphic to $\mathbb{Z}$, $\mathbb{Z}_2$ and $\mathbb{Z}_4$. He also proved that the image of $\mathcal{G}$ in $\mathcal{G}^{\mathbb{Q}}$ is isomorphic to a similar infinite direct sum. In the same article it is observed that $\mathcal{G}^{\mathbb{Q}}$ has a natural decomposition as a direct sum $\oplus\mathcal{G}^{\mathbb{Q}}_{p(t)}$, where the $p(t)$ are symmetric irreducible rational polynomials. We will not present the details here, but note that if the Alexander polynomial of K, $\Delta_K(t)$, is irreducible, then the image of K in $\mathcal{G}^{\mathbb{Q}}$ is in the $\mathcal{G}^{\mathbb{Q}}_{\Delta(t)}$ summand. Stoltzfus [1977] observed that the algebraic concordance group $\mathcal{G}$ does not have a similar splitting. Thus there is not an immediate analog in concordance for the decompositions we have been studying for homology cobordism. However he did prove that in some cases such a splitting exists. The following, from Corollary 6.5 of the same work, is stated in terms of knot concordance, but, given the isomorphism of higher-dimensional concordance and $\mathcal{G}^{\mathbb{Z}}$, the same splitting theorem holds in the algebraic concordance group.

Theorem 19. *If K is an n-dimensional knot for $n > 1$ and $\Delta_K(t)$ factors as $p(t)q(t)$ with $p(t)$ and $q(t)$ symmetric and the resultant* $\mathrm{Res}(p(t), q(t)) = 1$, *then K is concordant to a connected sum $K_1 \# K_2$, with $\Delta_{K_1}(t) = p(t)$ and $\Delta_{K_2}(t) = q(t)$.*

Here we observe that this result does not hold in dimension 3.

Example. Consider the ten-crossing knot $K = 10_5$. It has Alexander polynomial

$$\Delta = (1 - t + t^2)(1 - 2t + 2t^2 - t^3 + 2t^4 - 2t^5 + t^6).$$

These two factors are irreducible and have resultant 1.

Theorem 20. *The knot 10_5 is not concordant to any connected sum $K_1 \# K_2$, where $\Delta_{K_1} = 1 - t + t^2$ and $\Delta_{K_2} = 1 - 2t + 2t^2 - t^3 + 2t^4 - 2t^5 + t^6$.*

Proof. The 2-fold branched cover of K is the lens space $L(33, 13)$. If the desired concordance existed, then $L(33, 13)$ would split in rational cobordism as a connected sum $M_3 \# M_{11}$, with $H_1(M_3) = \mathbb{Z}_3$ and $H_1(M_{11}) = \mathbb{Z}_{11}$. In order to compute the relevant d-invariants, one first identifies $\mathfrak{s}_6$ as the spin structure $\mathfrak{s}_*$ by computing that the value of $d(L(33, 13), \mathfrak{s}_6) = 33$, a value that is not attained by any other Spin^c structure. The values of the d-invariants, $d(L(33, 13), (a, b)\cdot\mathfrak{s}_*) - d(L(33, 13), \mathfrak{s}_*)$ for $(a, b) \in \mathbb{Z}_3 \oplus \mathbb{Z}_{11}$ are given in the chart in Table 8 (multiplied by 33 to clear denominators).

The next chart, in Table 9, presents the values

$$\begin{aligned}\delta(L(33, 13), (a, b)) = d(L(33, 13), (a, b)) - d(L(33, 13), (a, 0))\\ - ad(L(33, 13), (0, b)) + d(L(33, 13), (0, 0)).\end{aligned}$$

a	$b = 0$	1	2	3	4	5
1	**22**	10	40	46	28	52
0	**0**	**54**	**18**	**24**	**6**	**30**
–1	**22**	10	40	–20	28	–14

Table 8. $33\, d(L(33, 13), 11a + 3b)$.

a	$b=0$	1	2	3	4	5
1	**0**	2	0	0	0	0
0	**0**	**0**	**0**	**0**	**0**	**0**
–1	**0**	2	0	2	0	2

Table 9. $\delta(L(33,13),(a,b))$.

The presence of the nonzero entries implies the nonsplittability of the manifold, as desired. □

Note. In [Livingston 2002], the second author constructed similar but much more complicated examples in the topological category.

13. Topologically trivial bordism

In [Hedden et al. 2012] the quotient $\Theta^T_{\mathbb{Q},\mathrm{spin}}/\Theta^I_{\mathbb{Q},\mathrm{spin}}$ was studied. Here the cobordism group has been restricted to spin 3-manifolds and spin bordisms which have the rational homology of S^3. The notation $\Theta^T_{\mathbb{Q},\mathrm{spin}}$ denotes the subgroup generated by representatives which bound topological homology balls and $\Theta^I_{\mathbb{Q},\mathrm{spin}}$ is generated by those that are cobordant to $\mathbb{Z}$-homology spheres. (Note that we have changed the notation from that of the same reference to be consistent with the results of the current paper.) There is also a similar result in the same work replacing $(\mathbb{Q},\mathrm{spin})$, with $\mathbb{Z}_2$. Recall that every orientable 3-manifold has trivial tangent bundle, so is spin, and that every $\mathbb{Z}_2$ homology 3-sphere has a unique spin structure.

Here we observe that Theorem 6 permits us to eliminate the need to constrain the cobordism group to being spin or to use $\mathbb{Z}_2$ coefficients. Let $\Theta^T_{\mathbb{Q}}$ denote the subgroup of $\Theta^3_{\mathbb{Q}}$ generated by rational homology spheres that are trivial in the topological rational cobordism group, that is, the kernel of $\mathcal{K}$.

Theorem 21. *The quotient group $\Theta^T_{\mathbb{Q}}/\Theta^3_{\mathbb{Z}}$ is infinitely generated.*

We outline how the argument in [Hedden et al. 2012] can be generalized.

In this work there is a family of rational homology spheres, M_{p^2}, constructed for an infinite set of primes p. These are constructed so that they bound topological balls. The proof of the theorem consists of showing that no linear combination $N = \#_i\, a_i M_{p_i^2} \# M_0$ bounds a spin rational homology ball (or $\mathbb{Z}_2$ homology ball) W, where M_0 is a $\mathbb{Z}$-homology sphere. The existence of a unique spin structure was used to identify Spin^c of the relevant manifolds with the second homology.

If all p are odd, then there is a unique Spin^c structure on N, and according to Theorem 6, it is the restriction of a Spin^c structure on W. Given this, Proposition 2.1 of the same reference, which required that W be spin, continues to apply to identify the Spin^c structures on N which extend to W with a metabolizer of the linking form on $H_1(N)$. That identification is what is used to obstruct the existence of W via d-invariants, as described in Theorem 3.2 of the same work. Thus the remainder of the proof goes through as in that paper.

Appendix A. Finding the p_i

The proof of Theorem 17 requires a sequence of odd pairs $\{p_i, p_i + 2\}$ such that the elements of the full set $\{p_i\} \cup \{p_i + 2\}$ are pairwise relatively prime and square-free. Since p_i and $p_i + 2$ are relatively prime, we need to choose the p_i so that all elements of $\{p_i(p_i + 2)\}$ are pairwise relatively prime and each element is square-free. If we let $p_i = n_i - 1$, then $p_i(p_i + 2) = n_i^2 - 1$, and so we are seeking an infinite sequence of positive integers $\{n_i\}$ such that:

(1) n_i is even for all i.

(2) All elements of $\{n_i^2 - 1\}$ are relatively prime.

(3) Each $n_i^2 - 1$ is square-free.

In Section 9 we need a sequence of integers n_i such that $n_i = 0 \mod 5$ with the property that the integers $20n_i^2 + 8n_i + 1$ are relatively prime and square-free. Here is a theorem that covers both cases.

Theorem 22. *Let $f(x) \in \mathbb{Z}[t]$ be an quadratic polynomial with constant term* 1 *that is not the square of a linear polynomial. Let α be a fixed integer and $s_n = \alpha n$ be an arithmetic sequence. There exists an infinite set of s_i such that values of $f(s_i)$ are pairwise relatively prime and square-free.*

Proof. It is known that if $g(n)$ is a quadratic polynomial that is not a square of a linear polynomial and which has the property that its coefficients have greatest common divisor one, then $g(n)$ is square-free for an infinite set of n (see, for example, [Erdős 1953]). We wish to construct the sequence of s_i inductively. To find s_1, let $f_1(n) = f(\alpha n)$, which is irreducible with constant term one. Choose n_1 so that $f_1(n_1)$ is square-free. Let $s_1 = \alpha n_1$. Assume that s_i has been defined for $i < k$. We find s_k with the desired properties as follows. Let $P = \prod_{i=1}^{k-1} f(s_i)$. Consider the function $f_k(n) = f(\alpha P n)$. Again this polynomial is irreducible with constant term one so there exists an n_k for which $f_k(n_k)$ is square-free. Since $f_k(n_k) = f(\alpha P n_k)$, we let $s_k = \alpha P n_k$. Notice that $f(\alpha P n) = 1 \mod p$ for each prime divisor p of P, since evaluating f at $\alpha P n$ gives a quadratic polynomial in n, with the quadratic term and linear term divisible by P and the constant term one. It follows that $f(s_k)$ is relatively prime to all $f(s_i), i < k$. □

Appendix B. The Alexander polynomial of $T_{p,p+2}$

Normalized to be symmetric, the Alexander polynomial of a knot can be written in the form $\Delta_K(t) = a_0 + \sum_{i=1}^{n} a_i(t^{-i} + t^i)$, where $a_0 + 2\sum a_i = \pm 1$. In Section 11 we use the following fact.

Theorem 23. *If $K = T_{p,p+2}$ with p odd then*

$$\Delta_{T_{p,p+2}}(t) = a_0 + \sum_{i=1}^{(p^2-1)/2} a_i(t^{-i}+t^i),$$

where $a_i = \pm 1$ for $i \leq (p-1)/2$.

Note. With more care, all the coefficients or $\Delta_{T_{p,p+2}}(t)$ can be described in closed form.

Proof. As a polynomial (as opposed to the normalized Laurent polynomial) with nonzero constant term, the Alexander polynomial of $T_{p,q}$ is

$$(1-t^{pq})(1-t)/(1-t^p)(1-t^q).$$

Expanding each term of the denominator in a power series and noting that multiplying by the t^{pq} term in the numerators does not affect terms of the product of degree less than $2g = (p-1)(q-1)$, the degree of the Alexander polynomial, we can focus on the expression

$$(1-t)(1+t^p+t^{2p}+t^{3p}+\cdots)(1+t^q+t^{2q}+\cdots),$$

which we write as the product

$$(1-t)\sum_{i=0}^{\infty} b_i t^i.$$

Here b_i is the number of solutions to $xp + yq = i$, with $x, y \geq 0$. In the case of interest, $q = p+2$ and the genus $g = (p^2-1)/2$. We will now show that for i in the range $g - A \leq i \leq g$, the values b_i are alternately 0 and 1, where A is a constant to be determined. Thus, using the fact that the Alexander polynomial is symmetric, upon multiplying by $(1-t)$ we have the coefficients of the Alexander polynomial are all ± 1 near t^g. To show that the coefficients b_i alternate between 0 and 1 for $g - A \leq i \leq g$, we first observe that in a given range of i, all $b_i \geq 1$ for i even. To see this, write $p = 2n+1$ and $q = 2n+3$; thus $g = 2n^2 + 2n$. Consider the sum

$$\frac{n+j}{2}p + \frac{n-j}{2}q = 2n^2 + 2n - j,$$

where j is selected to have the same parity as n. (We require here that $j \leq n$; that is, we need $A \leq (p-1)/2$.) To complete the argument, we next observe that the difference $|b_i - b_j| \leq 1$ if $|i-j| \leq 1$. Suppose otherwise. That is, suppose that there are *distinct* nonnegative solutions to the equations

$$xp + yq = i$$

and

$$x'p + y'q = j,$$

with $x, y, x', y' \geq 0$, $|i - j| \leq 1$, and $i, j \leq g$. The conditions that $i \leq g$ and $y \geq 0$ imply that $xp \leq g = (pq - p - q - 1)/2$, which implies that $x < (q-1)/2$. We first consider the case that $i \neq j$. After possibly reordering, the difference would give

$$(x - x')p + (y - y')q = 1.$$

One solution to this equation is

$$\frac{q-1}{2}p - \frac{p-1}{2}q = 1.$$

Every other solution is given by adding a multiple of $(-q, p)$ to the coefficient vector (note that $-q(p) + p(q) = 0$ is a primitive solution since p and q are relatively prime). Thus the solutions with the smallest absolute values of the x-coordinate to the unital equation are the one above and

$$-\frac{q+1}{2}p + \frac{p+1}{2}q = 1.$$

That is, the smallest possible value for $(x - x')$ is $x - x' = (q-1)/2$. But, since x and x' both are nonnegative and less than $(q-1)/2$, this is impossible. As an example, if $p = 21$ and $q = 23$ (so $g = 220$), we have the solutions

$$11(21) - 10(23) = 1$$

and

$$-12(21) + 11(23) = 1,$$

with $g = 220$. We also have $x(21) + y(23) \leq 220$ which implies that $x \leq 220/21$, so $0 \leq x \leq 10$. Similarly for x', so it is not possible for $|x - x'| = 11$. Finally we consider the case $i = j$. Thus our coefficients would satisfy

$$(x - x')p + (y - y')q = 0.$$

This implies that $x - x'$ is a multiple of q. But this would imply that they are equal since under our assumptions both are nonnegative and also $xp \leq pq - p - q + 1 \leq pq$, so $x < q$ and $x' < q$.

In summary, if we write the Alexander polynomial of the $T_{p,q}$ torus knot, with $q - p = 2$ as ± 1 as $a_0 + \sum_{i=1}^{g} a_i(t^i + t^{-i})$, then for $i \leq (p-1)/2$, we have shown that $a_i = (-1)^i$. $\square$

Acknowledgements

We are grateful for Matt Hedden's help in better understanding Heegaard–Floer homology. His results regarding the Heegaard–Floer theory of doubled knots are central here, and our specific examples are inspired by those that Matt pointed us toward in our collaborations with him. We also thank the referee for offering significant improvement in the exposition.

References

[Alexander et al. 1976] J. P. Alexander, G. C. Hamrick, and J. W. Vick, "Linking forms and maps of odd prime order", *Trans. Amer. Math. Soc.* **221**:1 (1976), 169–185. MR 53 #6600 Zbl 0357.57009

[Casson and Gordon 1986] A. J. Casson and C. M. Gordon, "Cobordism of classical knots", pp. 181–199 in *À la recherche de la topologie perdue*, edited by L. Guillou and A. Marin, Progr. Math. **62**, Birkhäuser, Boston, 1986. MR 900252 Zbl 0597.57001

[Erdős 1953] P. Erdős, "Arithmetical properties of polynomials", *J. London Math. Soc.* **28** (1953), 416–425. MR 15,104f Zbl 0051.27703

[Freedman 1982] M. H. Freedman, "The topology of four-dimensional manifolds", *J. Differential Geom.* **17**:3 (1982), 357–453. MR 84b:57006 Zbl 0528.57011

[Freedman and Quinn 1990] M. H. Freedman and F. Quinn, *Topology of* 4*-manifolds*, Princeton Mathematical Series **39**, Princeton University Press, 1990. MR 94b:57021 Zbl 0705.57001

[Furuta 1990] M. Furuta, "Homology cobordism group of homology 3-spheres", *Invent. Math.* **100**:2 (1990), 339–355. MR 91c:57039 Zbl 0716.55008

[Gompf and Stipsicz 1999] R. E. Gompf and A. I. Stipsicz, 4*-manifolds and Kirby calculus*, Graduate Studies in Mathematics **20**, American Mathematical Society, Providence, RI, 1999. MR 2000h:57038 Zbl 0933.57020

[Hancock et al. 2013] S. Hancock, J. Hom, and M. Newman, "On the knot Floer filtration of the concordance group", *J. Knot Theory Ramifications* **22**:14 (2013), Article ID #1350084. MR 3190122 Zbl 06273023

[Hedden 2007] M. Hedden, "Knot Floer homology of Whitehead doubles", *Geom. Topol.* **11** (2007), 2277–2338. MR 2008m:57030 Zbl 1187.57015

[Hedden et al. 2012] M. Hedden, C. Livingston, and D. Ruberman, "Topologically slice knots with nontrivial Alexander polynomial", *Adv. Math.* **231**:2 (2012), 913–939. MR 2955197 Zbl 1254.57008

[Hedden et al. 2014] M. Hedden, S. Kim, and C. Livingston, "Topologically slice knots of smooth concordance order two", preprint, 2014. arXiv 1212.6628

[Kawauchi and Kojima 1980] A. Kawauchi and S. Kojima, "Algebraic classification of linking pairings on 3-manifolds", *Math. Ann.* **253**:1 (1980), 29–42. MR 82b:57007 Zbl 0427.57001

[Levine 1969] J. Levine, "Invariants of knot cobordism", *Invent. Math.* **8** (1969), 98–110. MR 40 #6563 Zbl 0179.52401

[Lisca 2007] P. Lisca, "Sums of lens spaces bounding rational balls", *Algebr. Geom. Topol.* **7** (2007), 2141–2164. MR 2008m:57018 Zbl 1185.57015

[Livingston 2002] C. Livingston, "Examples in concordance", preprint, 2002. arXiv math/0101035v2

[Livingston and Naik 1999] C. Livingston and S. Naik, "Obstructing four-torsion in the classical knot concordance group", *J. Differential Geom.* **51**:1 (1999), 1–12. MR 2000g:57009 Zbl 1025.57013

[Milnor and Husemoller 1973] J. Milnor and D. Husemoller, *Symmetric bilinear forms*, Ergebnisse der Mathematik und ihrer Grenzgebiete **73**, Springer, New York, 1973. MR 58 #22129 Zbl 0292.10016

[Ozsváth and Szabó 2003] P. S. Ozsváth and Z. Szabó, "Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary", *Adv. Math.* **173**:2 (2003), 179–261. MR 2003m:57066 Zbl 1025.57016

[Ozsváth and Szabó 2004] P. S. Ozsváth and Z. Szabó, "Holomorphic disks and knot invariants", *Adv. Math.* **186**:1 (2004), 58–116. MR 2005e:57044 Zbl 1062.57019

[Ozsváth and Szabó 2005] P. S. Ozsváth and Z. Szabó, "On knot Floer homology and lens space surgeries", *Topology* **44**:6 (2005), 1281–1300. MR 2006f:57034 Zbl 1077.57012

[Stoltzfus 1977] N. W. Stoltzfus, *Unraveling the integral knot concordance group*, Mem. Amer. Math. Soc. **192**, American Mathematical Society, Providence, RI, 1977. MR 57 #7616 Zbl 0366.57005

Received June 17, 2013. Revised March 21, 2014.

SE-GOO KIM
DEPARTMENT OF MATHEMATICS AND RESEARCH INSTITUTE FOR BASIC SCIENCES
KYUNG HEE UNIVERSITY
SEOUL 130-701
SOUTH KOREA
sgkim@khu.ac.kr

CHARLES LIVINGSTON
DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
RAWLES HALL
BLOOMINGTON, IN 47405-5701
UNITED STATES
livingst@indiana.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.213

BIHARMONIC SURFACES OF CONSTANT MEAN CURVATURE

ERIC LOUBEAU AND CEZAR ONICIUC

We compute a Simons-type formula for the stress-energy tensor of biharmonic maps from surfaces. Specializing to Riemannian immersions, we prove several rigidity results for biharmonic CMC surfaces, putting in evidence the influence of the Gaussian curvature on pseudoumbilicity. Finally the condition of biharmonicity is shown to enable an extension of the classical Hopf theorem to CMC surfaces in any ambient Riemannian manifold.

1. Introduction

While harmonic maps between abstract Riemannian manifolds are a generalization of minimal submanifolds, their study on two-dimensional domains remained nonetheless very valuable and brought new light to both theories. When, for topological or geometrical reasons, harmonic maps are nonexistent or unsatisfactory, one can then measure the failure of harmonicity with the *bienergy functional*

$$E_2(\phi) = \frac{1}{2}\int_M |\tau(\phi)|^2 \, v_g,$$

where M is compact, $\phi : (M, g) \to (N, h)$ is a smooth map and $\tau(\phi) = \operatorname{trace} \nabla d\phi$ is the tension field. Usual arguments (see [Jiang 1986]) show that critical points of E_2, called *biharmonic maps*, are solutions of

$$\tau_2(\phi) = -\Delta\tau(\phi) - \operatorname{trace} R^N(d\phi(\cdot), \tau(\phi))d\phi(\cdot) = 0,$$

and we will use the adjective *proper* to designate nonharmonic biharmonic maps.

Whilst the interconnections between harmonic maps and minimal surfaces are clear and well-established, in many cases, but not always, biharmonic Riemannian immersions have constant mean curvature (CMC). However, this link is not as clear as harmonicity and minimality, and the principal objective of this article is to explain how biharmonicity constrains CMC surfaces in an abstract ambient

C. Oniciuc was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PN-II-RU-TE-2011-3-0108.
MSC2010: 53C42, 53C43, 58E20.
Keywords: biharmonic maps, constant mean curvature, stress-energy tensor.

manifold. This is particularly well-illustrated on compact biharmonic CMC surfaces whose Gaussian curvature has constant sign. They must be flat or pseudoumbilical if K^M is nonnegative (Corollary 11); otherwise they have pseudoumbilical points (Theorem 12). The role of pseudoumbilical points in relaxing curvature constraints is further felt in the noncompact case, as their absence forces the CMC surface to be conformally flat (Theorem 12).

For complete surfaces, nonnegative Gaussian curvature and an upper bound on the sectional curvature of the ambient space will cause the surface to be flat or pseudoumbilical, but note that both can occur simultaneously (Proposition 18). When the ambient manifold is a three-dimensional space form, the surface must be umbilical (Corollary 14); consult [Montaldo and Oniciuc 2006] for the classification.

Our approach is to derive, in Proposition 3, a Simons-type formula for the biharmonic stress-energy tensor, valid for all smooth maps. As cumbersome as this equation is in the general case, on biharmonic maps from surfaces it simplifies enough (Proposition 5) to enable the use of a divergence argument (Theorem 6) and draw some consequences (Corollaries 8 and 9). However, the main consequences are for CMC biharmonic surfaces.

To close the article, we show that, in any ambient space, the condition of biharmonicity preserves the holomorphicity of the Hopf differential of CMC surfaces (Theorem 20).

Biharmonic CMC surfaces were also studied in [Fetcu and Pinheiro 2013; Ou and Wang 2011] and [Sasahara 2007].

The conventions we adopt are that the Riemann curvature tensor is

$$R(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]},$$

while its (0, 4) counterpart is

$$R(X, Y, Z, W) = \langle R(X, Y)W, Z\rangle.$$

The choice of sign for the Laplacians on sections and functions is the same, and on the real line $\Delta f = -f''$.

All objects, unless specified, are smooth and we assume summation on repeated indices, when apt.

2. The biharmonic stress-energy tensor on surfaces

Since biharmonic maps stem from a variational problem, one can apply the general principle of studying the same functional but under variations of the domain metric. This idea taken up on the bienergy leads to the biharmonic stress-energy tensor, which is symmetric and of type (0, 2); see [Loubeau et al. 2008].

Definition 1. Let (M, g) and (N, h) be Riemannian manifolds and $\phi : M \to N$ a smooth map. The biharmonic stress-energy tensor of ϕ is

$$S_2(X, Y) = \left\{ \frac{|\tau(\phi)|^2}{2} + \langle d\phi, \nabla\tau(\phi)\rangle \right\} g(X, Y) - T(X, Y),$$

where $T(X, Y) = \langle d\phi(X), \nabla_Y \tau(\phi)\rangle + \langle d\phi(Y), \nabla_X \tau(\phi)\rangle$.

The main feature of S_2 is satisfying Hilbert's principle of being divergence-free at critical points [Loubeau et al. 2008; Jiang 1987]; that is, $\operatorname{div} S_2 = -\langle d\phi, \tau_2(\phi)\rangle$.

In order to exploit the biharmonicity of the map ϕ, we compute the rough Laplacian of its biharmonic stress-energy tensor. This second-order operator on $(0, 2)$-tensors will reveal curvature terms which combine with the bitension field, and formulas will involve swapping vector positions in the third fundamental form of ϕ, with curvature appearing according to a lemma we quote separately, without proof.

Lemma 2. *Let $\phi : (M^m, g) \to (N^n, h)$ be a smooth map. Then*

$$(\nabla^2 d\phi)(X, Y, Z) - (\nabla^2 d\phi)(Z, Y, X) = R(X, Z)d\phi(Y) - d\phi(R^M(X, Z)Y)$$

for any $X, Y, Z \in C(TM)$.

Proposition 3 (the rough Laplacian of S_2). *Let (M, g) and (N, h) be Riemannian manifolds and $\phi : M \to N$ a smooth map; then the (rough) Laplacian of S_2 is the symmetric $(0, 2)$-tensor*

$$\begin{aligned}
&(\Delta^R S_2)(X, Y)\\
&\quad= \big(\langle \Delta\tau(\phi), \tau(\phi)\rangle - 2|\nabla\tau(\phi)|^2 - 2\textstyle\sum\langle R(X_i, X_j)d\phi(X_i), \nabla_{X_j}\tau(\phi)\rangle\\
&\qquad\qquad - 2\langle d\phi(\mathrm{Ricci}^M(\,\cdot\,)), \nabla_{(\cdot)}\tau(\phi)\rangle - 2\langle \nabla d\phi, \nabla^2\tau(\phi)\rangle + \langle d\phi, \nabla(\Delta\tau(\phi))\rangle\\
&\qquad\qquad - \langle \nabla(\operatorname{trace} R^N(d\phi(\,\cdot\,), \tau(\phi))d\phi(\,\cdot\,)), d\phi\rangle\\
&\qquad\qquad - \langle \operatorname{trace} R^N(d\phi(\,\cdot\,), \tau(\phi))d\phi(\,\cdot\,), \tau(\phi)\rangle\big) g(X, Y)\\
&\qquad + 2\langle \nabla_X\tau(\phi), \nabla_Y\tau(\phi)\rangle + \textstyle\sum\langle R(X_i, X)d\phi(X_i), \nabla_Y\tau(\phi)\rangle\\
&\qquad + \textstyle\sum\langle R(X_i, Y)d\phi(X_i), \nabla_X\tau(\phi)\rangle + \langle d\phi(\mathrm{Ricci}^M(X)), \nabla_Y\tau(\phi)\rangle\\
&\qquad + \langle d\phi(\mathrm{Ricci}^M(Y)), \nabla_X\tau(\phi)\rangle + 2\textstyle\sum\langle \nabla d\phi(X_i, X), (\nabla^2\tau(\phi))(X_i, Y)\rangle\\
&\qquad + 2\textstyle\sum\langle \nabla d\phi(X_i, Y), (\nabla^2\tau(\phi))(X_i, X)\rangle - \langle d\phi(X), \nabla_Y(\Delta\tau(\phi))\rangle\\
&\qquad - \langle d\phi(Y), \nabla_X(\Delta\tau(\phi))\rangle + \textstyle\sum\langle d\phi(X), R(X_i, Y)\nabla_{X_i}\tau(\phi)\rangle\\
&\qquad + \textstyle\sum\langle d\phi(Y), R(X_i, X)\nabla_{X_i}\tau(\phi)\rangle + \textstyle\sum\langle d\phi(X), \nabla_{X_i}R(X_i, Y)\tau(\phi)\rangle\\
&\qquad + \textstyle\sum\langle d\phi(Y), \nabla_{X_i}R(X_i, X)\tau(\phi)\rangle + \langle d\phi(X), \nabla_{\mathrm{Ricci}^M(Y)}\tau(\phi)\rangle\\
&\qquad + \langle d\phi(Y), \nabla_{\mathrm{Ricci}^M(X)}\tau(\phi)\rangle,
\end{aligned}$$

where $\{X_i\}$ is a geodesic frame around the point $p \in M$.

Proof. Let $\phi : (M, g) \to (N, h)$ be a smooth map between Riemannian manifolds. We will work with a geodesic frame $\{X_i\}$ around the point $p \in M$ and evaluate at p.

Writing out the Laplacian in the geodesic frame yields

$$\begin{aligned}
&\Delta(\langle d\phi, \nabla\tau(\phi)\rangle) \\
&\quad = -\sum\big\{\langle \nabla_{X_i}[\nabla d\phi(X_i, X_j)], \nabla_{X_j}\tau(\phi)\rangle + 2\langle \nabla d\phi(X_i, X_j), (\nabla^2\tau(\phi))(X_i, X_j)\rangle \\
&\qquad\qquad + \langle d\phi(X_j), \nabla_{X_i}\nabla_{X_i}\nabla_{X_j}\tau(\phi)\rangle - \langle d\phi(X_j), \nabla_{X_i}\nabla_{\nabla_{X_i}X_j}\tau(\phi)\rangle\big\},
\end{aligned}$$

and by the symmetry formula of the third fundamental form, we have

$$\sum \nabla_{X_i}[\nabla d\phi(X_i, X_j)] = \nabla_{X_j}\tau(\phi) + \sum R(X_i, X_j) d\phi(X_i) + d\phi(\mathrm{Ricci}^M(X_j)),$$

and

$$\begin{aligned}
&\sum\big(\nabla_{X_i}\nabla_{X_i}\nabla_{X_j}\tau(\phi) - \nabla_{X_i}\nabla_{\nabla_{X_i}X_j}\tau(\phi)\big) \\
&\qquad = \sum\big\{\nabla_{X_j}\nabla_{X_i}\nabla_{X_i}\tau(\phi) + \nabla_{[X_i, X_j]}\nabla_{X_i}\tau(\phi) + R(X_i, X_j)\nabla_{X_i}\tau(\phi) \\
&\qquad\qquad + \nabla_{X_i}R(X_i, X_j)\tau(\phi) \\
&\qquad\qquad - \big(\nabla_{\nabla_{X_j}X_i}\nabla_{X_i}\tau(\phi) + \nabla_{[X_i, \nabla_{X_j}X_i]}\tau(\phi) + R(X_i, \nabla_{X_j}X_i)\tau(\phi)\big)\big\} \\
&\qquad = -\nabla_{X_j}\big(\Delta\tau(\phi)\big) + \sum\big\{(\nabla^2\tau(\phi))(X_j, \nabla_{X_i}X_i) + R(X_i, X_j)\nabla_{X_i}\tau(\phi) \\
&\qquad\qquad\qquad + \nabla_{X_i}R(X_i, X_j)\tau(\phi)\big\} \\
&\qquad\qquad + \nabla_{\mathrm{Ricci}^M(X_j)}\tau(\phi),
\end{aligned}$$

since

$$\sum[X_i, \nabla_{X_j}X_i] = \sum \nabla_{X_j}\nabla_{X_i}X_i - \mathrm{Ricci}^M(X_j).$$

Therefore

$$\begin{aligned}
&\Delta\big(\langle d\phi, \nabla\tau(\phi)\rangle\big) \\
&\quad = -\sum\langle \nabla_{X_j}\tau(\phi), \nabla_{X_j}\tau(\phi)\rangle - \sum\langle R(X_i, X_j) d\phi(X_i), \nabla_{X_j}\tau(\phi)\rangle \\
&\qquad - \sum\langle d\phi(\mathrm{Ricci}^M(X_j)), \nabla_{X_j}\tau(\phi)\rangle \\
&\qquad - 2\langle \nabla d\phi, \nabla^2\tau(\phi)\rangle + \sum\langle d\phi(X_j), \nabla_{X_j}(\Delta\tau(\phi))\rangle \\
&\qquad - \sum\big(\langle d\phi(X_j), R(X_i, X_j)\nabla_{X_i}(\tau(\phi))\rangle + \langle d\phi(X_j), \nabla_{X_i}R(X_i, X_j)\tau(\phi)\rangle \\
&\qquad\qquad + \langle d\phi(X_j), \nabla_{\mathrm{Ricci}^M(X_j)}\tau(\phi)\rangle\big),
\end{aligned}$$

but

$$\begin{aligned}
&\sum\langle d\phi(X_j), \nabla_{X_i}R(X_i, X_j)\tau(\phi)\rangle \\
&\qquad = \sum X_i\langle \operatorname{trace} R^N(d\phi(\cdot), \tau(\phi)) d\phi(\cdot), d\phi(X_i)\rangle - \langle \nabla d\phi, R(\cdot, \cdot)\tau(\phi)\rangle \\
&\qquad = \langle \nabla(\operatorname{trace} R^N(d\phi(\cdot), \tau(\phi)) d\phi(\cdot)), d\phi\rangle \\
&\qquad\qquad + \langle \operatorname{trace} R^N(d\phi(\cdot), \tau(\phi)) d\phi(\cdot), \tau(\phi)\rangle,
\end{aligned}$$

and

$$\sum\langle R(X_i, X_j)d\phi(X_i), \nabla_{X_j}\tau(\phi)\rangle = \sum\langle R(X_i, X_j)\nabla_{X_i}\tau(\phi), d\phi(X_j)\rangle,$$

whilst

$$\sum\langle d\phi(\mathrm{Ricci}^M(\,\cdot\,)), \nabla\tau(\phi)\rangle = \sum\langle d\phi(\,\cdot\,), \nabla_{\mathrm{Ricci}^M(\,\cdot\,)}\tau(\phi)\rangle,$$

so the Laplacian of the scalar term is

$$\begin{aligned}\Delta&\Big(\frac{|\tau(\phi)|^2}{2} + \langle d\phi, \nabla\tau(\phi)\rangle\Big)\\ &= \langle \Delta\tau(\phi), \tau(\phi)\rangle - 2|\nabla\tau(\phi)|^2\\ &\quad - 2\sum\langle R(X_i, X_j)d\phi(X_i), \nabla_{X_j}\tau(\phi)\rangle - 2\langle d\phi(\mathrm{Ricci}^M(\,\cdot\,)), \nabla_{(\cdot)}\tau(\phi)\rangle\\ &\quad - 2\langle \nabla d\phi, \nabla^2\tau(\phi)\rangle + \langle d\phi(\,\cdot\,), \nabla_{(\cdot)}(\Delta\tau(\phi))\rangle\\ &\quad - \langle \nabla(\mathrm{trace}\, R^N(d\phi(\,\cdot\,), \tau(\phi))d\phi(\,\cdot\,)), d\phi\rangle\\ &\quad - \langle \mathrm{trace}\, R^N(d\phi(\,\cdot\,), \tau(\phi))d\phi(\,\cdot\,), \tau(\phi)\rangle.\end{aligned}$$

On the other hand, to compute the (rough) Laplacian of the symmetric two-tensor

$$T(X, Y) = \langle d\phi(X), \nabla_Y\tau(\phi)\rangle + \langle d\phi(Y), \nabla_X\tau(\phi)\rangle,$$

we put $X = X_k$ and $Y = X_j$ and obtain, still evaluating expressions at the point p,

$$\begin{aligned}-(\Delta^R T)(X, Y) = \sum\big(&\langle \nabla_{X_i}\nabla_{X_i}d\phi(X), \nabla_Y\tau(\phi)\rangle + 2\langle \nabla_{X_i}d\phi(X), \nabla_{X_i}\nabla_Y\tau(\phi)\rangle\\ &+ \langle d\phi(X), \nabla_{X_i}\nabla_{X_i}\nabla_Y\tau(\phi)\rangle + \langle \nabla_{X_i}\nabla_{X_i}d\phi(Y), \nabla_X\tau(\phi)\rangle\\ &+ 2\langle \nabla_{X_i}d\phi(Y), \nabla_{X_i}\nabla_X\tau(\phi)\rangle + \langle d\phi(Y), \nabla_{X_i}\nabla_{X_i}\nabla_X\tau(\phi)\rangle\\ &- \langle d\phi(Y), \nabla_{X_i}\nabla_{\nabla_{X_i}X}\tau(\phi)\rangle - \langle d\phi(X), \nabla_{X_i}\nabla_{\nabla_{X_i}Y}\tau(\phi)\rangle\big),\end{aligned}$$

since $\nabla_{X_i}\nabla_{X_i}X_j$ vanishes at the point p. This last expression simplifies further if we use the symmetries properties of the third fundamental form of ϕ to obtain

$$\begin{aligned}\sum \nabla_{X_i}\nabla_{X_i}d\phi(X) &= \sum(\nabla^2 d\phi)(X_i, X_i, X)\\ &= \nabla_X\tau(\phi) + \sum R(X_i, X)d\phi(X_i) + d\phi(\mathrm{Ricci}^M(X)),\end{aligned}$$

and the curvature tensor of the pullback bundle for

$$\begin{aligned}\sum\big(&\langle d\phi(X), \nabla_{X_i}\nabla_{X_i}\nabla_Y\tau(\phi)\rangle - \langle d\phi(X), \nabla_{X_i}\nabla_{\nabla_{X_i}Y}\tau(\phi)\rangle\big)\\ &= -\langle d\phi(X), \nabla_Y(\Delta\tau(\phi))\rangle + \sum\langle d\phi(X), R(X_i, Y)\nabla_{X_i}\tau(\phi)\rangle\\ &\quad + \sum\langle d\phi(X), \nabla_{X_i}R(X_i, Y)\tau(\phi)\rangle + \langle d\phi(X), \nabla_{\mathrm{Ricci}^M(Y)}\tau(\phi)\rangle.\end{aligned}$$

The Laplacian of the tensor T is then equal to

$$\begin{aligned}
-(\Delta^R T)(X, Y)& \\
=&\langle \nabla_X \tau(\phi), \nabla_Y \tau(\phi)\rangle + \sum \langle R(X_i, X) d\phi(X_i), \nabla_Y \tau(\phi)\rangle \\
&+ \langle d\phi(\mathrm{Ricci}^M(X)), \nabla_Y \tau(\phi)\rangle + 2\sum \langle \nabla d\phi(X_i, X), (\nabla^2 \tau(\phi))(X_i, Y)\rangle \\
&- \langle d\phi(X), \nabla_Y(\Delta\tau(\phi))\rangle + \sum \langle d\phi(X), R(X_i, Y)\nabla_{X_i}\tau(\phi)\rangle \\
&+ \sum \langle d\phi(X), \nabla_{X_i} R(X_i, Y)\tau(\phi)\rangle + \langle d\phi(X), \nabla_{\mathrm{Ricci}^M(Y)}\tau(\phi)\rangle \\
&+ \langle \nabla_Y \tau(\phi), \nabla_X \tau(\phi)\rangle + \sum \langle R(X_i, Y) d\phi(X_i), \nabla_X \tau(\phi)\rangle \\
&+ \langle d\phi(\mathrm{Ricci}^M(Y)), \nabla_X \tau(\phi)\rangle + 2\sum \langle \nabla d\phi(X_i, Y), (\nabla^2 \tau(\phi))(X_i, X)\rangle \\
&- \langle d\phi(Y), \nabla_X(\Delta\tau(\phi))\rangle + \sum \langle d\phi(Y), R(X_i, X)\nabla_{X_i}\tau(\phi)\rangle \\
&+ \sum \langle d\phi(Y), \nabla_{X_i} R(X_i, X)\tau(\phi)\rangle + \langle d\phi(Y), \nabla_{\mathrm{Ricci}^M(X)}\tau(\phi)\rangle.
\end{aligned}$$

Summing the various parts together yields the proposition. □

Remark 4. In order to see the geometric meaning of the term $\sum \nabla_{X_i} R(X_i, X)\tau(\phi)$, we can rewrite it as $\sum (\nabla R)(X_i, X_i, X, \tau(\phi)) + \sum R(X_i, X)\nabla_{X_i}\tau(\phi)$.

While the general expression for the rough Laplacian of S_2 at first seems unwieldy, in a manner reminiscent of its harmonic counterpart (see [Baird et al. 2011]) it becomes amenable when the domain is a surface and the map biharmonic. The final formula only involves three ingredients: the tensor S_2 itself, the Gaussian curvature and the norm of the tension field of the map. This paves the way for a series of propositions and corollaries for both maps and Riemannian immersions, where topological and curvature conditions restrict the existence of biharmonic maps.

Proposition 5. *Let $\phi : (M^2, g) \to (N, h)$ be a biharmonic map defined on a surface M^2. The Laplacian of its biharmonic stress-energy tensor is*

$$\Delta^R S_2 = -2K^M S_2 + \nabla d(|\tau(\phi)|^2) + \{K^M |\tau(\phi)|^2 + \Delta|\tau(\phi)|^2\} g,$$

where K^M is the Gaussian curvature of (M^2, g).

Proof. Since $\dim M = 2$, its Ricci curvature is $\mathrm{Ricci}^M = K^M I$, with $K^M \in C^\infty(M)$. We will work with a geodesic frame $\{X_1, X_2\}$ around a point $p \in M^2$ and evaluate all expressions at this point.

As $\Delta^R S_2$ is a symmetric $(0, 2)$-tensor, there are only two cases to consider, and, from the previous proposition, combined with basic symmetries of the curvature tensor and the biharmonicity condition, we have

$$
\begin{aligned}
&(\Delta^R S_2)(X_1, X_2)\\
&= 2\langle \nabla_{X_1}\tau(\phi), \nabla_{X_2}\tau(\phi)\rangle - \langle d\phi(X_2), \nabla_{X_1}(\Delta\tau(\phi))\rangle\\
&\quad + 2K^M\{\langle d\phi(X_1), \nabla_{X_2}\tau(\phi)\rangle + \langle d\phi(X_2), \nabla_{X_1}\tau(\phi)\rangle\}\\
&\quad + 2\langle \nabla d\phi(X_1, X_2), -\Delta\tau(\phi)\rangle + 2\langle \nabla d\phi(X_1, X_1), (\nabla^2\tau(\phi))(X_1, X_2)\rangle\\
&\quad + 2\langle \nabla d\phi(X_2, X_2), (\nabla^2\tau(\phi))(X_2, X_1)\rangle - \langle d\phi(X_1), \nabla_{X_2}(\Delta\tau(\phi))\rangle\\
&\quad + \langle d\phi(X_1), \nabla_{X_1} R(X_1, X_2)\tau(\phi)\rangle + \langle d\phi(X_2), \nabla_{X_2} R(X_2, X_1)\tau(\phi)\rangle\\
&= 2\langle \nabla_{X_1}\tau(\phi), \nabla_{X_2}\tau(\phi)\rangle + 2K^M\{\langle d\phi(X_1), \nabla_{X_2}\tau(\phi)\rangle + \langle d\phi(X_2), \nabla_{X_1}\tau(\phi)\rangle\}\\
&\quad + 2\langle \nabla d\phi(X_1, X_1), (\nabla^2\tau(\phi))(X_1, X_2)\rangle + 2\langle \nabla d\phi(X_2, X_2), (\nabla^2\tau(\phi))(X_2, X_1)\rangle\\
&\quad - \langle \nabla_{X_1} d\phi(X_1), R(X_1, X_2)\tau(\phi)\rangle - \langle \nabla_{X_2} d\phi(X_2), R(X_2, X_1)\tau(\phi)\rangle.
\end{aligned}
$$

But

$$
\begin{aligned}
&2\langle \nabla d\phi(X_1, X_1), (\nabla^2\tau(\phi))(X_1, X_2)\rangle - \langle \nabla d\phi(X_1, X_1), R(X_1, X_2)\tau(\phi)\rangle\\
&\qquad = \langle \nabla d\phi(X_1, X_1), 2\nabla_{X_1}\nabla_{X_2}\tau(\phi) - \nabla_{X_1}\nabla_{X_2}\tau(\phi) + \nabla_{X_2}\nabla_{X_1}\tau(\phi)\rangle,
\end{aligned}
$$

so

$$
\begin{aligned}
(\Delta^R S_2)(X_1, X_2) &= 2K^M\{\langle d\phi(X_1), \nabla_{X_2}\tau(\phi)\rangle + \langle d\phi(X_2), \nabla_{X_1}\tau(\phi)\rangle\}\\
&\quad + 2\langle \nabla_{X_1}\tau(\phi), \nabla_{X_2}\tau(\phi)\rangle + \langle \tau(\phi), \nabla_{X_1}\nabla_{X_2}\tau(\phi) + \nabla_{X_2}\nabla_{X_1}\tau(\phi)\rangle.
\end{aligned}
$$

Since

$$
\begin{aligned}
&(\nabla d|\tau(\phi)|^2)(X_1, X_2)\\
&\qquad = \langle \nabla_{X_1}\nabla_{X_2}\tau(\phi) + \nabla_{X_2}\nabla_{X_1}\tau(\phi), \tau(\phi)\rangle + 2\langle \nabla_{X_1}\tau(\phi), \nabla_{X_2}\tau(\phi)\rangle,
\end{aligned}
$$

we deduce that

$$(\Delta^R S_2)(X_1, X_2) = -2K^M S_2(X_1, X_2) + (\nabla d|\tau(\phi)|^2)(X_1, X_2).$$

The other case to look at is when the two vectors are the same, and then Proposition 3 shows that, using the symmetries of R^N,

$$
\begin{aligned}
&(\Delta^R S_2)(X_1, X_1)\\
&= -2\langle R^N(d\phi(X_1), \tau(\phi))d\phi(X_1), \tau(\phi)\rangle\\
&\quad -2\langle R^N(d\phi(X_2), \tau(\phi))d\phi(X_2), \tau(\phi)\rangle - 2\langle \nabla_{X_2}\tau(\phi), \nabla_{X_2}\tau(\phi)\rangle\\
&\quad -2K^M\langle d\phi(X_2), \nabla_{X_2}\tau(\phi)\rangle - 2\langle \nabla d\phi(X_2, X_2), (\nabla^2\tau(\phi))(X_2, X_2)\rangle\\
&\quad -2\langle \nabla d\phi(X_1, X_2), (\nabla^2\tau(\phi))(X_1, X_2)\rangle + 2\langle d\phi(X_1), \nabla_{X_1}(\Delta\tau(\phi))\rangle\\
&\quad +2\langle d\phi(X_2), \nabla_{X_2}(\Delta\tau(\phi))\rangle + 2K^M\langle d\phi(X_1), \nabla_{X_1}\tau(\phi)\rangle\\
&\quad +2\langle \nabla d\phi(X_1, X_1), (\nabla^2\tau(\phi))(X_1, X_1)\rangle + 2\langle \nabla d\phi(X_2, X_1), (\nabla^2\tau(\phi))(X_2, X_1)\rangle\\
&\quad -2\langle d\phi(X_1), \nabla_{X_1}(\Delta\tau(\phi))\rangle + 2\langle d\phi(X_1), \nabla_{X_2} R(X_2, X_1)\tau(\phi)\rangle
\end{aligned}
$$

$$
\begin{aligned}
&= -2|\nabla_{X_2}\tau(\phi)|^2 - 2K^M\langle d\phi(X_2), \nabla_{X_2}\tau(\phi)\rangle + 2K^M\langle d\phi(X_1), \nabla_{X_1}\tau(\phi)\rangle \\
&\quad - 2\langle R^N(d\phi(X_1), \tau(\phi))d\phi(X_1), \nabla d\phi(X_1, X_1)\rangle \\
&\quad - 2X_2\langle d\phi(X_2), R^N(d\phi(X_1), \tau(\phi))d\phi(X_1)\rangle \\
&\quad - 2\langle R^N(d\phi(X_2), \tau(\phi))d\phi(X_2), \nabla d\phi(X_1, X_1)\rangle \\
&\quad - 2\langle \nabla d\phi(X_2, X_2), (\nabla^2\tau(\phi))(X_2, X_2)\rangle \\
&\quad + 2\langle \nabla d\phi(X_1, X_1), (\nabla^2\tau(\phi))(X_1, X_1)\rangle - 2\langle \nabla d\phi(X_1, X_2), R(X_1, X_2)\tau(\phi)\rangle \\
&\quad + 2\langle d\phi(X_1), \nabla_{X_2}R(X_2, X_1)\tau(\phi)\rangle,
\end{aligned}
$$

since

i) $$
\begin{aligned}
&-2\langle \nabla d\phi(X_1, X_2), (\nabla^2\tau(\phi))(X_1, X_2) \\
&\quad +2\langle \nabla d\phi(X_2, X_1), (\nabla^2\tau(\phi))(X_2, X_1)\rangle \\
&\qquad = -2\langle \nabla d\phi(X_1, X_2), R(X_1, X_2)\tau(\phi)\rangle,
\end{aligned}
$$

ii) $$
\begin{aligned}
&-2\langle R^N(d\phi(X_1), \tau(\phi))d\phi(X_1), \tau(\phi)\rangle \\
&\quad -2\langle d\phi(X_2), \nabla_{X_2}R^N(d\phi(X_1), \tau(\phi))d\phi(X_1)\rangle \\
&\qquad = -2\langle R^N(d\phi(X_1), \tau(\phi))d\phi(X_1), \nabla_{X_1}d\phi(X_1)\rangle \\
&\qquad\quad - 2\langle R^N(d\phi(X_1), \tau(\phi))d\phi(X_1), \nabla_{X_2}d\phi(X_2)\rangle \\
&\qquad\quad - 2X_2\langle d\phi(X_2), R^N(d\phi(X_1), \tau(\phi))d\phi(X_1)\rangle \\
&\qquad\quad + 2\langle \nabla_{X_2}d\phi(X_2), R^N(d\phi(X_1), \tau(\phi))d\phi(X_1)\rangle,
\end{aligned}
$$

iii) $$
\begin{aligned}
&-2\langle R^N(d\phi(X_2), \tau(\phi))d\phi(X_2), \tau(\phi)\rangle \\
&\quad -2\langle d\phi(X_2), \nabla_{X_2}R^N(d\phi(X_2), \tau(\phi))d\phi(X_2)\rangle \\
&\qquad = -2\langle R^N(d\phi(X_2), \tau(\phi))d\phi(X_2), \nabla d\phi(X_1, X_1)\rangle.
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&-X_2\langle d\phi(X_2), R^N(d\phi(X_1), \tau(\phi))d\phi(X_1)\rangle + \langle d\phi(X_1), \nabla_{X_2}R(X_2, X_1)\tau(\phi)\rangle \\
&= -X_2R^N(d\phi(X_2), d\phi(X_1), d\phi(X_1), \tau(\phi)) \\
&\quad + X_2R^N(d\phi(X_2), d\phi(X_1), d\phi(X_1), \tau(\phi)) + \langle \nabla d\phi(X_1, X_2), R(X_1, X_2)\tau(\phi)\rangle
\end{aligned}
$$

so

$$
\begin{aligned}
&(\Delta^R S_2)(X_1, X_1) \\
&= -2|\nabla_{X_2}\tau(\phi)|^2 - 2K^M\langle d\phi(X_2), \nabla_{X_2}\tau(\phi)\rangle \\
&\quad + 2K^M\langle d\phi(X_1), \nabla_{X_1}\tau(\phi)\rangle - 2\langle R^N(d\phi(X_1), \tau(\phi))d\phi(X_1), \nabla d\phi(X_1, X_1)\rangle \\
&\quad + 2\langle \nabla d\phi(X_1, X_1), (\nabla^2\tau(\phi))(X_1, X_1)\rangle \\
&\quad - 2\langle R^N(d\phi(X_2), \tau(\phi))d\phi(X_2), \nabla d\phi(X_1, X_1)\rangle \\
&\quad - 2\langle \nabla d\phi(X_2, X_2), (\nabla^2\tau(\phi))(X_2, X_2)\rangle.
\end{aligned}
$$

But

$$\langle \tau(\phi), \nabla_{X_2}\nabla_{X_2}\tau(\phi)\rangle = -\tfrac{1}{2}\Delta|\tau(\phi)|^2 - \tfrac{1}{2}X_1X_1(|\tau(\phi)|^2) - |\nabla_{X_2}\tau(\phi)|^2,$$

so

$$\begin{aligned}(\Delta^R S_2)(X_1, X_1) = -2K^M S_2(X_1, X_1) &+ \big\{K^M|\tau(\phi)|^2 + \Delta|\tau(\phi)|^2\big\}g(X_1, X_1)\\ &+ (\nabla d|\tau(\phi)|^2)(X_1, X_1),\end{aligned}$$

with a similar expression for $(\Delta^R S_2)(X_2, X_2)$. Therefore

$$\Delta^R S_2 = -2K^M S_2 + \nabla d(|\tau(\phi)|^2) + \{K^M|\tau(\phi)|^2 + \Delta|\tau(\phi)|^2\}g. \qquad \square$$

The expression for the Laplacian of the biharmonic stress-energy tensor on a surface is simple enough to be contracted with S_2 itself and combined with the divergence theorem, if the domain is assumed to be compact. The ensuing integral formula tightly binds the tensor S_2, the Gaussian curvature and the norm of the tension field together, and conditions on two of them determine the third.

More geometrical applications will be found for Riemannian immersions in the next section.

Theorem 6. *Let $\phi : M^2 \to N^n$ be a biharmonic map and assume M^2 is compact. Then*

$$\int_M |\nabla S_2|^2\, v_g + 2\int_M K^M\left(|S_2|^2 - \frac{|\tau(\phi)|^4}{2}\right) v_g = \int_M |d(|\tau(\phi)|^2)|^2\, v_g,$$

where K^M is the Gaussian curvature of (M^2, g).

Proof. Observe that

$$\operatorname{div}\langle S_2, d(|\tau(\phi)|^2)\rangle = \langle \operatorname{div} S_2, d(|\tau(\phi)|^2)\rangle + \langle S_2, \operatorname{Hess}(|\tau(\phi)|^2)\rangle.$$

As $\operatorname{div} S_2 = 0$, we have

$$\int_M \langle S_2, \operatorname{Hess}(|\tau(\phi)|^2)\rangle\, v_g = \int_M \operatorname{div}\{\langle S_2, d(|\tau(\phi)|^2)\rangle\}\, v_g = 0,$$

which combined with the classical equality

$$\int_M \langle \Delta^R S_2, S_2\rangle\, v_g = \int_M |\nabla S_2|^2\, v_g$$

gives the theorem. $\square$

Remark 7. Note that the term $2|S_2|^2 - |\tau(\phi)|^4$ is always nonnegative since it is equal to $(S_2(X_1, X_1) - S_2(X_2, X_2))^2 + 4S_2^2(X_1, X_2)$, and $|S_2|^2 = |\tau(\phi)|^4/2$ if and only if $S_2 = |\tau(\phi)|^2 g/2$.

A biharmonic map with parallel stress-energy tensor must have a tension field of constant norm [Loubeau et al. 2008], but Proposition 5 shows greater restrictions for two-dimensional domains.

Corollary 8. *Let $\phi : M^2 \to N^n$ be a biharmonic map, and assume M is compact and $\nabla S_2 = 0$. Then $|\tau(\phi)|$ is constant and $\int_M K^M \, v_g = 0$ or $S_2 = |\tau(\phi)|^2 g/2$.*

Proof. If $\nabla S_2 = 0$, then its norm and trace, $|\tau(\phi)|^2$, are constant, hence

$$\left(|S_2|^2 - \frac{|\tau(\phi)|^4}{2}\right) \int_M K^M \, v_g = 0. \qquad \square$$

If the norm of the tension field is constant, we can deduce a partial converse for nonnegative curvature.

Corollary 9. *Let $\phi : (M^2, g) \to (N^n, h)$ be a proper -biharmonic map with $|\tau(\phi)|^2$ constant. Assume M is compact and $K^M \geq 0$. Then S_2 is parallel and M is flat or $S_2 = |\tau(\phi)|^2 g/2$.*

3. Constant mean curvature surfaces

To be able to offer conditions with greater geometrical content, we concentrate our applications on Riemannian immersions. The recurrent condition on the map is pseudoumbilicity, as an equality between the shape operator A_H in the direction of the mean curvature vector field H and the metric.

The pivotal role of pseudoumbilical immersions, already observed in the study of the biharmonic stress-energy tensor (see [Loubeau et al. 2008]), emerges again in connection with the curvature of the domain surface, sometimes to the extent of determining its topology.

In the absence of compactness, the divergence theorem is substituted with a parabolicity argument on constant mean curvature surfaces, associated with a bound on the curvature tensor of the target space.

Finally, working with complex coordinates on a Riemann surface, the $(2, 0)$-part of the H-component of the second fundamental form B is shown to be holomorphic if and only if the mean curvature is constant.

Recall that if $\phi : M^2 \to N$ is a pseudoumbilical proper-biharmonic Riemannian immersion then it is CMC. As a consequence, and since $S_2 = -2|H|^2 g + 4A_H$, a rewording of Corollaries 8 and 9 is as follows:

Corollary 10. *Let $\phi : (M^2, g) \to (N^n, h)$ be a proper-biharmonic Riemannian immersion from a compact oriented surface, with $\nabla A_H = 0$. Then M is topologically a torus or pseudoumbilical.*

Corollary 11. *Let $\phi : (M^2, g) \to (N^n, h)$ be a CMC proper-biharmonic Riemannian immersion. Assume M is compact and $K^M \geq 0$. Then $\nabla A_H = 0$ and M is flat or pseudoumbilical.*

The next result shows that pseudoumbilical points allow some flexibility of the curvature; since away from these points special coordinates exist in which the metric is conformally flat (with a globally defined factor), the shape operator has a simple expression, while its eigenvalues can be computed from the mean curvature vector field (see [Hasanis and Vlachos 1996] for a similar result).

Theorem 12. *Let $\phi : (M^2, g) \to (N^n, h)$ be a CMC proper-biharmonic Riemannian immersion. We denote by λ_1 and λ_2 the principal curvatures of M corresponding to A_H, with $\lambda_1 \geq \lambda_2$, and let $\mu = \lambda_1 - \lambda_2$. Consider $p \in M$ such that $\mu(p) > 0$; that is, p is a nonpseudoumbilical point. Then, around p there is a local chart $(U; x, y)$ which is both isothermal and a line of curvature coordinate system for A_H. We have, on U,*

$$g = \frac{1}{\mu}(dx^2 + dy^2), \qquad \langle A_H(\,\cdot\,), \cdot\,\rangle = \frac{1}{\mu}(\lambda_1\, dx^2 + \lambda_2\, dy^2),$$
$$\sum_{i=1}^{2} R^N(X_i, H, X_i, H) - |\nabla^\perp H|^2 - 2|H|^4 > 0,$$

and

$$\lambda_1 = |H|^2 + \frac{\sqrt{2}}{2}\sqrt{\textstyle\sum_{i=1}^{2} R^N(X_i, H, X_i, H) - |\nabla^\perp H|^2 - 2|H|^4}\,,$$
$$\lambda_2 = |H|^2 - \frac{\sqrt{2}}{2}\sqrt{\textstyle\sum_{i=1}^{2} R^N(X_i, H, X_i, H) - |\nabla^\perp H|^2 - 2|H|^4}\,,$$

with $X_1 = \sqrt{\mu}\, \partial x$, $X_2 = \sqrt{\mu}\, \partial y$. Moreover

$$\Delta \ln\biggl(\sum_{i=1}^{2} R^N(X_i, H, X_i, H) - |\nabla^\perp H|^2 - 2|H|^4\biggr) = -4K^M,$$

and, in codimension one, the Gauss equation becomes

$$\mathrm{Riem}^N(X_1, X_2) = K^M - 2|H|^2 + \frac{1}{2|H|^2}\,\mathrm{Ricci}^N(H, H).$$

Proof. Let λ_1 and λ_2 be the principal curvatures in the direction of H; that is, λ_1 and λ_2 are the eigenvalues of A_H. In an open neighborhood U around a nonpseudoumbilical point p, $\lambda_1 > \lambda_2$ on U and $\lambda_1, \lambda_2 \in C^\infty(U)$ (in general they are only continuous), and therefore $\mu = \lambda_1 - \lambda_2$ is a positive smooth function on U.

Let $\{X_1, X_2\}$ be a local orthonormal frame on U such that $A_H(X_1) = \lambda_1 X_1$ and $A_H(X_2) = \lambda_2 X_2$. We consider $\omega_1^2, \omega_2^1 \in \bigwedge^1(U)$ defined by

$$\nabla X_1 = \omega_1^2 X_2 \quad \text{and} \quad \nabla X_2 = \omega_2^1 X_1.$$

Clearly $\omega_1^2 = -\omega_2^1$. If we put $X = Z = X_1$ and $Y = X_2$, the Codazzi equation becomes

$$R^N(X_1, H, X_2, X_1) \\ = -\omega_2^1(X_1)\mu - X_2\lambda_1 - \langle B(X_2, X_1), \nabla^\perp_{X_1} H\rangle + \langle B(X_1, X_1), \nabla^\perp_{X_2} H\rangle.$$

Recall that, since $|H|$ is constant, the tangent part of the biharmonic equation is

$$\operatorname{trace} A_{\nabla^\perp_{\cdot} H}(\,\cdot\,) + \operatorname{trace}(R^N(d\phi(\,\cdot\,), H)d\phi(\,\cdot\,))^T = 0.$$

Taking the inner product with X_2, we have

$$\langle B(X_2, X_1), \nabla^\perp_{X_1} H\rangle + \langle B(X_2, X_2), \nabla^\perp_{X_2} H\rangle + R^N(X_1, H, X_2, X_1) = 0;$$

thus

$$\omega_2^1(X_1)\mu + X_2\lambda_1 = 2\langle H, \nabla^\perp_{X_2} H\rangle = 0$$

and

$$\omega_2^1(X_1) = -\frac{X_2\lambda_1}{\mu}.$$

Note that

$$X_2(\lambda_2) = X_2\langle A_H(X_2), X_2\rangle = X_2\langle B(X_2, X_2), H\rangle = -X_2\lambda_1,$$

therefore

$$\omega_2^1(X_1) = \frac{1}{2}\left(-\frac{X_2\lambda_1}{\mu} + \frac{X_2\lambda_2}{\mu}\right) = -\frac{1}{2}\frac{X_2\mu}{\mu}.$$

Exchanging X_1 and X_2, we similarly obtain

$$\omega_2^1(X_2) = \frac{1}{2}\frac{X_1\mu}{\mu},$$

therefore

$$\omega_2^1 = -\frac{1}{2}\frac{X_2\mu}{\mu}\omega_1 + \frac{1}{2}\frac{X_1\mu}{\mu}\omega_2.$$

The Gauss equation implies that

$$d\omega_2^1(X_1, X_2) = K^M;$$

that is,

$$K^M = \tfrac{1}{2}(X_1X_1 \ln\mu + X_2X_2 \ln\mu) - (\omega_2^1(X_1))^2 - (\omega_2^1(X_2))^2,$$

but

$$\nabla_{X_1} X_1 = \tfrac{1}{2}(X_2 \ln \mu) X_2, \qquad (\nabla_{X_1} X_1)(\ln \mu) = \tfrac{1}{2}(X_2 \ln \mu)^2,$$
$$(\omega_2^1(X_1))^2 = \tfrac{1}{4}(X_2 \ln \mu)^2 = \tfrac{1}{2}(\nabla_{X_1} X_1)(\ln \mu),$$

while

$$\nabla_{X_2} X_2 = \tfrac{1}{2}(X_1 \ln \mu) X_1, \qquad (\nabla_{X_2} X_2)(\ln \mu) = \tfrac{1}{2}(X_1 \ln \mu)^2,$$
$$(\omega_2^1(X_2))^2 = \tfrac{1}{4}(X_1 \ln \mu)^2 = \tfrac{1}{2}(\nabla_{X_2} X_2)(\ln \mu).$$

Therefore

$$\Delta \ln \mu = -2K^M.$$

Since

$$\left[\frac{1}{\sqrt{\mu}} X_1, \frac{1}{\sqrt{\mu}} X_2\right] = 0,$$

there exist coordinate functions (x, y) on U such that $\partial/\partial x = X_1/\sqrt{\mu}$ and $\partial/\partial y = X_2/\sqrt{\mu}$. Moreover, the normal part of the biharmonicity equation

$$\Delta^\perp H + \operatorname{trace} B(\,\cdot\,, A_H \,\cdot\,) + \operatorname{trace}(R^N(\,\cdot\,, H)\,\cdot\,)^\perp = 0$$

implies, when H is constant,

$$|\nabla^\perp H|^2 + |A_H|^2 - \sum_{i=1}^{2} R^N(X_i, H, X_i, H) = 0,$$

and, since

$$\lambda_1 + \lambda_2 = 2|H|^2 \quad \text{and} \quad \lambda_1^2 + \lambda_2^2 = |A_H|^2,$$

we deduce that

$$|A_H|^2 - 2|H|^4 = \frac{(\lambda_1 - \lambda_2)^2}{2},$$

hence

$$\lambda_1 - \lambda_2 = \sqrt{2}\sqrt{\textstyle\sum_{i=1}^{2} R^N(X_i, H, X_i, H) - |\nabla^\perp H|^2 - 2|H|^4}. \qquad \square$$

Remark 13. i) If $n = 3$, we can replace $\sum_{i=1}^{2} R^N(X_i, H, X_i, H)$ by $\operatorname{Ricci}^N(H, H)$.
ii) Let $\phi : (M^2, g) \to (N, h)$ be a CMC proper-biharmonic Riemannian immersion. If (M^2, g) is complete and has no pseudoumbilical point then its universal cover is (globally) conformally equivalent to $\mathbb{R}^2$.

Corollary 14. *Let $\phi : (M^2, g) \to N^3(c)$ be a CMC proper-biharmonic Riemannian immersion in a three-dimensional real space form. Then it is umbilical.*

Proof. If there exists a nonumbilical point $p_0 \in M$, then, around p_0, we have

$$\operatorname{Riem}^N(X_1, X_2) = K^M - 2|H|^2 + \frac{1}{2|H|^2}\operatorname{Ricci}^N(H, H)$$

and

$$K^M = -\tfrac{1}{4}\Delta \ln(\operatorname{Ricci}^N(H, H) - 2|H|^4),$$

but $\operatorname{Ricci}^N(H, H) = 2c|H|^2$ is constant, so K^M is zero. On the other hand, the first equation implies that $c = K^M - 2|H|^2 + c$, which contradicts $K^M = 0$. □

As the formulas for λ_1 and λ_2 in Theorem 12 remain valid also for pseudoumbilical points, we deduce:

Corollary 15. *Let $\phi : (M^2, g) \to (N^3, h)$ be a CMC proper-biharmonic Riemannian immersion. Assume that there exists $c > 0$ such that $\operatorname{Ricci}^N(U, U) \geq c|U|^2$ with $|H|^2 \in (0, c/2)$. Then M^2 has no pseudoumbilical point.*

Corollary 16. *Let $\phi : (M^2, g) \to (N^n, h)$ be a CMC proper-biharmonic Riemannian immersion. Assume M is compact, oriented and has no pseudoumbilical point; then M is topologically a torus.*

Corollary 17. *Let $\phi : (M^2, g) \to (N^n, h)$ be a proper-biharmonic Riemannian immersion. Assume that λ_1 and λ_2 are constant; then $\nabla A_H = 0$, and M is flat or pseudoumbilical.*

If M is not compact, we need some assumption on the curvature of the target space (see also [Fetcu and Pinheiro 2013, Proposition 4.6 and 4.7]).

Proposition 18. *Let $\phi : (M^2, g) \to (N^n, h)$ be a CMC proper-biharmonic Riemannian immersion. Assume M is noncompact and complete and K^M is nonnegative. Assume that $\operatorname{Riem}^N \leq K_0$, where $K_0 > 0$ (in the sense that $R^N(U, V, U, V) \leq K_0$ for all $\{U, V\}$ orthonormal). Then $\nabla A_H = 0$, and M is flat or pseudoumbilical.*

Proof. By the previous formulas for the Laplacian of S_2, we have

$$\begin{aligned} -\tfrac{1}{2}\Delta|S_2|^2 &= -\langle \Delta^R S_2, S_2\rangle + |\nabla S_2|^2 \\ &= K^M(2|S_2|^2 - |\tau(\phi)|^4) + |\nabla S_2|^2, \end{aligned}$$

which must be nonnegative (Remark 7); therefore $|S_2|^2$ is a subharmonic function and bounded from above since, for Riemannian immersions,

$$|S_2|^2 = 8(2|A_H|^2 - 3|H|^4)$$

and $|A_H|^2$ is itself bounded from above. Indeed if ϕ is biharmonic, then

$$\Delta^\perp H + \operatorname{trace} B(\cdot, A_H \cdot) + \operatorname{trace}(R^N(\cdot, H)\cdot)^\perp = 0;$$

thus

$$
\begin{aligned}
|A_H|^2 &= -|\nabla^\perp H|^2 + \sum_{i=1}^{2} R^N(X_i, H, X_i, H) \\
&\leq \sum_{i=1}^{2} R^N(X_i, H, X_i, H) \leq 2|H|^2 K_0,
\end{aligned}
$$

and $|A_H|^2 \leq 2K_0|H|^2$. As M is complete with K^M nonnegative, it is parabolic and $|S_2|^2$, a subharmonic function bounded from above, must be constant (see [Huber 1957]):

$$K^M(|A_H|^2 - 4|H|^4) = 0,$$

while $\nabla A_H = 0$; in particular, $|A_H|^2$ is constant. □

Remark 19. When the dimension of the target is three, we can replace the curvature condition by an upper bound on the Ricci tensor.

The Hopf theorem [1983] shows that a compact simply connected CMC surface immersed in a three-dimensional Euclidean space must be umbilical, hence an embedded round sphere, and the condition of biharmonicity allows us to extend this to any codomain. This result has some strict implications on the set of pseudoumbilical points and hints at the difficulties of working with non-CMC surfaces. An interesting parallel can be drawn with [Fetcu and Pinheiro 2013].

Theorem 20. *Let $\phi : (M^2, g) \to (N^n, h)$ be a proper-biharmonic Riemannian immersion with mean curvature vector field H, with M^2 oriented. Let z be a complex coordinate on M^2; then the function $\langle B(\partial z, \partial z), H\rangle$ is holomorphic if and only if the norm of H is constant.*

Proof. The tangent part of the biharmonic equation is

$$\operatorname{grad} \frac{|H|^2}{2} + \operatorname{trace} A_{\nabla^\perp_\cdot H}(\,\cdot\,) + \operatorname{trace}(R^N(d\phi(\,\cdot\,), H)d\phi(\,\cdot\,))^T = 0.$$

Let $g = \lambda^2(dx^2 + dy^2)$ and

$$
\begin{aligned}
&\tfrac{1}{2}\partial x(|H|^2)\partial x + \tfrac{1}{2}\partial y(|H|^2)\partial y + A_{\nabla^\perp_{\partial x} H}(\partial x) + A_{\nabla^\perp_{\partial y} H}(\partial y) \\
&\qquad\qquad + (R^N(\partial x, H)\partial x + R^N(\partial y, H)\partial y)^T = 0;
\end{aligned}
$$

therefore

$$\frac{\lambda^2}{2}\partial x(|H|^2) + \langle A_{\nabla^\perp_{\partial x} H}(\partial x), \partial x\rangle + \langle A_{\nabla^\perp_{\partial y} H}(\partial y), \partial x\rangle + R^N(\partial y, H, \partial x, \partial y) = 0,$$

and

$$\frac{\lambda^2}{2}\partial y(|H|^2) + \langle A_{\nabla^\perp_{\partial x} H}(\partial x), \partial y\rangle + \langle A_{\nabla^\perp_{\partial y} H}(\partial y), \partial y\rangle + R^N(\partial x, H, \partial y, \partial x) = 0,$$

which is equivalent to

$$\text{(1)}\quad \frac{\lambda^2}{2}\partial x(|H|^2)+\langle B(\partial x,\partial x),\nabla^\perp_{\partial x}H\rangle+\langle B(\partial x,\partial y),\nabla^\perp_{\partial y}H\rangle \\ +R^N(\partial y,H,\partial x,\partial y)=0,$$

and

$$\text{(2)}\quad \frac{\lambda^2}{2}\partial y(|H|^2)+\langle B(\partial y,\partial x),\nabla^\perp_{\partial x}H\rangle+\langle B(\partial y,\partial y),\nabla^\perp_{\partial y}H\rangle \\ +R^N(\partial x,H,\partial y,\partial x)=0.$$

Since $\partial z=(\partial x-i\partial y)/2$ and $\partial\bar z=(\partial x+i\partial y)/2$, we see that

$$B(\partial z,\partial z)=\tfrac12(\lambda^2H-B(\partial y,\partial y)-iB(\partial x,\partial y))$$

and

$$\langle B(\partial z,\partial z),H\rangle=\tfrac12(\lambda^2|H|^2-\langle B(\partial y,\partial y),H\rangle-i\langle B(\partial x,\partial y),H\rangle).$$

Next we compute $\partial\bar z\langle B(\partial z,\partial z),H\rangle$:

$$\begin{aligned}
&(\partial x+i\partial y)\big(\lambda^2|H|^2-\langle B(\partial y,\partial y),H\rangle-i\langle B(\partial x,\partial y),H\rangle\big)\\
&\ =2\lambda\frac{\partial\lambda}{\partial x}|H|^2+\lambda^2\partial x(|H|^2)-\langle\nabla^\perp_{\partial x}B(\partial y,\partial y),H\rangle-\langle B(\partial y,\partial y),\nabla^\perp_{\partial x}H\rangle\\
&\qquad+\langle\nabla^\perp_{\partial y}B(\partial x,\partial y),H\rangle+\langle B(\partial x,\partial y),\nabla^\perp_{\partial y}H\rangle\\
&\qquad+i\bigg\{2\lambda\frac{\partial\lambda}{\partial y}|H|^2+\lambda^2\partial y(|H|^2)-\langle\nabla^\perp_{\partial y}B(\partial y,\partial y),H\rangle-\langle B(\partial y,\partial y),\nabla^\perp_{\partial y}H\rangle\\
&\qquad\qquad-\langle\nabla^\perp_{\partial x}B(\partial x,\partial y),H\rangle-\langle B(\partial x,\partial y),\nabla^\perp_{\partial x}H\rangle\bigg\}\\
&\ =A+iB.
\end{aligned}$$

With (1),

$$\begin{aligned}
A&=2\lambda\frac{\partial\lambda}{\partial x}|H|^2+\tfrac12\lambda^2\partial x(|H|^2)-\langle\nabla^\perp_{\partial x}B(\partial y,\partial y),H\rangle-\langle B(\partial y,\partial y),\nabla^\perp_{\partial x}H\rangle\\
&\qquad+\langle\nabla^\perp_{\partial y}B(\partial x,\partial y),H\rangle-\langle B(\partial x,\partial x),\nabla^\perp_{\partial x}H\rangle-R(\partial y,H,\partial x,\partial y)\\
&=2\lambda\frac{\partial\lambda}{\partial x}|H|^2+\tfrac12\lambda^2\partial x(|H|^2)-\langle\nabla^\perp_{\partial x}B(\partial y,\partial y),H\rangle+\langle\nabla^\perp_{\partial y}B(\partial x,\partial y),H\rangle\\
&\qquad-\langle 2\lambda^2H,\nabla^\perp_{\partial x}H\rangle-R(\partial y,H,\partial x,\partial y)\\
&=2\lambda\frac{\partial\lambda}{\partial x}|H|^2-\tfrac12\lambda^2\partial x(|H|^2)-\langle\nabla^\perp_{\partial x}B(\partial y,\partial y),H\rangle+\langle\nabla^\perp_{\partial y}B(\partial x,\partial y),H\rangle\\
&\qquad-R(\partial y,H,\partial x,\partial y).
\end{aligned}$$

From the Codazzi equation,

$$\begin{aligned}
&\langle \nabla^{\perp}_{\partial y} B(\partial x, \partial y), H\rangle \\
&\qquad = \langle (\nabla^{\perp}_{\partial y} B)(\partial x, \partial y), H\rangle + \langle B(\nabla_{\partial y}\partial x, \partial y), H\rangle + \langle B(\partial x, \nabla_{\partial y}\partial y), H\rangle \\
&\qquad = \langle \nabla^{\perp}_{\partial x} B(\partial y, \partial y), H\rangle - 2\langle B(\nabla_{\partial x}\partial y, \partial y), H\rangle + R(\partial y, \partial x, H, \partial y) \\
&\qquad\qquad + \langle B(\nabla_{\partial y}\partial x, \partial y), H\rangle + \langle B(\partial x, \nabla_{\partial y}\partial y), H\rangle;
\end{aligned}$$

therefore

$$\begin{aligned}
A &= 2\lambda \frac{\partial\lambda}{\partial x}|H|^2 - \tfrac{1}{2}\lambda^2 \partial x(|H|^2) - \langle \nabla^{\perp}_{\partial x} B(\partial y, \partial y), H\rangle + \langle \nabla^{\perp}_{\partial x} B(\partial y, \partial y), H\rangle \\
&\qquad - \langle B(\nabla_{\partial x}\partial y, \partial y), H\rangle + \langle B(\partial x, \nabla_{\partial y}\partial y), H\rangle + R(\partial y, \partial x, H, \partial y) \\
&\qquad - R(\partial y, H, \partial x, \partial y) \\
&= 2\lambda \frac{\partial\lambda}{\partial x}|H|^2 - \tfrac{1}{2}\lambda^2 \partial x(|H|^2) - \left\langle B\left(\frac{1}{\lambda}\left(\frac{\partial\lambda}{\partial y}\partial x + \frac{\partial\lambda}{\partial x}\partial y\right), \partial y\right), H\right\rangle \\
&\qquad + \left\langle B\left(\frac{1}{\lambda}\left(-\frac{\partial\lambda}{\partial x}\partial x + \frac{\partial\lambda}{\partial y}\partial y\right), \partial x\right), H\right\rangle \\
&= -\tfrac{1}{2}\lambda^2 \partial x(|H|^2).
\end{aligned}$$

Identical arguments for the imaginary part B, using (2), yield

$$B = \tfrac{1}{2}\lambda^2 \partial y(|H|^2). \qquad \square$$

Remark 21. If $\phi : (M^2, g) \to (N^n, h)$ is a CMC proper-biharmonic Riemannian immersion, with M^2 oriented. Then $\langle B(\partial z, \partial z), H\rangle dz^2$ is globally defined and, if M^2 has no pseudoumbilical point, it is equal to $dz^2/4$ and therefore M^2 is an affine manifold.

Corollary 22. *Let $\phi : (M^2, g) \to (N^n, h)$ be a CMC proper-biharmonic Riemannian immersion, with M^2 oriented. If M^2 is not pseudoumbilical, then its pseudoumbilical points are isolated.*

Theorem 20 yields:

Theorem 23. *Let $\phi : (M^2, g) \to (N^n, h)$ be a CMC proper-biharmonic Riemannian immersion. If M^2 is a topological sphere $\mathbb{S}^2$, then M is pseudoumbilical.*

Proof. Since $\langle B(\partial z, \partial z), H\rangle = 0$, we have

$$\langle B(\partial x, \partial x) - B(\partial y, \partial y), H\rangle = 0 \quad \text{and} \quad \langle B(\partial x, \partial y), H\rangle = 0,$$

which is equivalent to

$$\langle A_H(\partial x), \partial x\rangle = \langle A_H(\partial y), \partial y\rangle \quad \text{and} \quad \langle A_H(\partial x), \partial y\rangle = \langle A_H(\partial y), \partial x\rangle = 0. \quad \square$$

References

[Baird et al. 2011] P. Baird, E. Loubeau, and C. Oniciuc, "Harmonic and biharmonic maps from surfaces", pp. 223–230 in *Harmonic maps and differential geometry* (Cagliari, 2009), edited by E. Loubeau and S. Montaldo, Contemp. Math. **542**, Amer. Math. Soc., Providence, RI, 2011. MR 2012e:53121 Zbl 1236.58024

[Fetcu and Pinheiro 2013] D. Fetcu and A. L. Pinheiro, "Biharmonic surfaces with parallel mean curvature in complex space forms", preprint, 2013. arXiv 1303.4279v1

[Hasanis and Vlachos 1996] T. Hasanis and T. Vlachos, "2-type surfaces in a hypersphere", *Kodai Math. J.* **19**:1 (1996), 26–38. MR 96m:53059 Zbl 0856.53019

[Hopf 1983] H. Hopf, *Differential geometry in the large*, Lecture Notes in Math. **1000**, Springer, Berlin, 1983. MR 85b:53001 Zbl 0526.53002

[Huber 1957] A. Huber, "On subharmonic functions and differential geometry in the large", *Comment. Math. Helv.* **32** (1957), 13–72. MR 20 #970 Zbl 0080.15001

[Jiang 1986] G. Y. Jiang, "2-harmonic maps and their first and second variational formulas", *Chinese Ann. Math.* (*A*) **7**:4 (1986), 389–402. In Chinese; summary translated in *Chinese Ann. Math.* (*B*) **7**:4 (1986), p. 523. MR 88i:58039 Zbl 0628.58008

[Jiang 1987] G. Y. Jiang, "The conservation law for 2-harmonic maps between Riemannian manifolds", *Acta Math. Sinica* **30**:2 (1987), 220–225. In Chinese. MR 88k:58028 Zbl 0631.58007

[Loubeau et al. 2008] E. Loubeau, S. Montaldo, and C. Oniciuc, "The stress-energy tensor for biharmonic maps", *Math. Z.* **259**:3 (2008), 503–524. MR 2009c:58020 Zbl 1139.58010

[Montaldo and Oniciuc 2006] S. Montaldo and C. Oniciuc, "A short survey on biharmonic maps between Riemannian manifolds", *Rev. Un. Mat. Argentina* **47**:2 (2006), 1–22. MR 2008a:53063 Zbl 1140.58004

[Ou and Wang 2011] Y.-L. Ou and Z.-P. Wang, "Constant mean curvature and totally umbilical biharmonic surfaces in 3-dimensional geometries", *J. Geom. Phys.* **61**:10 (2011), 1845–1853. MR 2012g:53120 Zbl 1227.58004

[Sasahara 2007] T. Sasahara, "Biharmonic Lagrangian surfaces of constant mean curvature in complex space forms", *Glasg. Math. J.* **49**:3 (2007), 497–507. MR 2008i:53081 Zbl 1132.53310

Received June 11, 2013.

ERIC LOUBEAU
DÉPARTEMENT DE MATHÉMATIQUES
UNIVERSITÉ DE BRETAGNE OCCIDENTALE
6, AVENUE VICTOR LE GORGEU
CS 93837
29238 BREST 3
FRANCE
loubeau@univ-brest.fr

CEZAR ONICIUC
FACULTY OF MATHEMATICS
ALEXANDRU IOAN CUZA UNIVERSITY OF IASI
BOULEVARD CAROL I, NUMBER 11
700506 IASI
ROMANIA
oniciucc@uaic.ro

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.231

FOLIATIONS OF A SMOOTH METRIC MEASURE SPACE BY HYPERSURFACES WITH CONSTANT f-MEAN CURVATURE

JUNCHEOL PYO

We study smooth codimension-one foliations $\mathscr{F}$ of a smooth metric measure space whose leaves have the same constant f-mean curvature. Firstly, we show that all the leaves of $\mathscr{F}$ are f-minimal hypersurfaces when either the smooth metric measure space is compact and has nonnegative Bakry–Émery Ricci curvature, or the limit of the ratio of the weighted volume of a geodesic ball B and the weighted area of a geodesic sphere ∂B vanishes. Secondly, we prove that every leaf of $\mathscr{F}$ is strongly f-stable. Lastly, we show that there is no complete proper foliation of the Gaussian space whose leaves have the same constant f-mean curvature. In particular, there are no foliations of $\mathbb{R}^{n+1}$ whose leaves are complete proper self-similar solutions for mean curvature flow.

1. Introduction and the statement of results

The study of smooth codimension-one foliations of manifolds has a long history in mathematics (see [Lawson 1974] and reference therein). In [Barbosa et al. 1987; 1991; Meeks 1988; Oshikiri 1981], there are very interesting results on foliations whose leaves have constant mean curvature. In this paper, we consider foliations of a smooth metric measure space whose leaves are hypersurfaces having the same f-mean curvatures. The main questions we consider here concern the rigidity and f-minimality of such foliations of a smooth metric measure space. Extending the classical results (i.e., when f is constant) to a smooth metric measure space requires f or $|\nabla f|$ to be bounded in many cases; see [Morgan 2005; Wei and Wylie 2009], for example. Our proof follows the one from the case where f is constant [Barbosa et al. 1987; 1991] but without any further assumption on f. Moreover, for particular weight functions f, we get rigidity results for self-similar surfaces or translating solitons which are models for singularities of mean curvature flow.

This work was supported by a Two-Year Research Grant from Pusan National University.
MSC2010: primary 53C12; secondary 53C42.
Keywords: foliation, constant f-mean curvature, f-stable, smooth metric measure space.

Recall that a smooth metric measure space $(M^{n+1}, \bar{g}, f)$ is a smooth Riemannian manifold $(M^{n+1}, \bar{g})$ with a positive density e^{-f} used to weight the volume of domains and the area of hypersurfaces. Let Σ be an isometrically immersed hypersurface in $(M^{n+1}, \bar{g})$. Denote by dv and dA the Riemannian volume forms on M and Σ with respect to $\bar{g}$ and the induced metric $g = i^*\bar{g}$, respectively. Then the weighted volume and area are given by $dv_m = e^{-f}\,dv$ and $dA_m = e^{-f}\,dA$, respectively.

Smooth metric measure spaces naturally arise in various fields. The Gaussian space, i.e., Euclidean space with the Gaussian density $e^{-\pi|x|^2}$, appears in the study of probability and statistics. Many interesting solitons in geometric flows (e.g., self-similar solutions and translating solitons to the mean curvature flow, and Ricci solitons to the Ricci flow) are represented by f-minimal hypersurfaces in a smooth metric measure space (see [Bakry and Émery 1985; Cheng et al. 2012; Colding and Minicozzi 2012; Huisken and Sinestrari 1999; Morgan 2005; Pyo 2014] and the references therein).

With the upper bar, we denote the geometric quantities on the ambient space $(M^{n+1}, \bar{g})$. For example, $\overline{\nabla}$, $\bar{d}$, $\overline{\nabla}^2$, $\bar{\triangle}$, $\overline{\mathrm{div}}$ and $\overline{\mathrm{Ric}}$, denote the Levi-Civita connection, exterior differentiation, Hessian, Laplacian, divergence and Ricci tensor of $(M^{n+1}, \bar{g})$, respectively. For a smooth metric measure space, we naturally consider the *Bakry–Émery Ricci tensor* $\overline{\mathrm{Ric}}_f$, which is defined by

$$\overline{\mathrm{Ric}}_f = \overline{\mathrm{Ric}} + \overline{\nabla}^2 f,$$

and the f-Laplacian $\bar{\triangle}_f = \bar{\triangle} - \bar{g}(\overline{\nabla} f, \overline{\nabla})$ on M, which is a selfadjoint operator with respect to the weighted measure dv_m. For a smooth vector field ξ, the f-divergence of ξ is defined by

$$\overline{\mathrm{div}}_f \xi = e^f\, \overline{\mathrm{div}}(e^{-f}\xi).$$

Let ν be a unit normal vector field to Σ in M. With the induced metric $g = i^*\bar{g}$ on Σ, the second fundamental form of (Σ, g) is given by $A(X, Y) = g(\overline{\nabla}_X Y, \nu)$ for any two tangent vectors X and Y on Σ, and the mean curvature by $H = \mathrm{tr}(A)$. For the hypersurface Σ in $(M, \bar{g}, f)$, we define the f-mean curvature H_f with respect to ν as follows:

$$H_f = H + \bar{g}(\overline{\nabla} f, \nu),$$

which is obtained by the first variation formula of the weighted area. For (Σ, g), ∇, d, $\triangle$ and div denote the Levi-Civita connection, exterior differentiation, Laplacian and divergence on Σ, respectively.

The following is proved for foliations of a compact smooth metric measure space with nonnegative Bakry–Émery Ricci curvature:

Theorem 2. *Let $(M^{n+1}, \bar{g}, f)$ be a compact smooth metric measure space with nonnegative Bakry–Émery Ricci curvature and $\mathcal{F}$ a codimension-one smooth foliation of M whose leaves have the same constant f-mean curvature. Then every leaf of $\mathcal{F}$ is a totally geodesic and f-minimal hypersurface with vanishing Bakry–Émery Ricci curvature in the normal direction.*

In a smooth metric measure space $(M^{n+1}, \bar{g}, f)$, we define the ratio

$$\Lambda_f(R, p) = \frac{\mathrm{vol}_f(\partial B_p(R))}{\overline{\mathrm{vol}}_f(B_p(R))},$$

where $\overline{\mathrm{vol}}_f(B_p(R))$ and $\mathrm{vol}_f(\partial B_p(R))$ are the weighted volume of the geodesic ball $B_p(R)$ and the geodesic sphere $\partial B_p(R)$ for a point p, respectively. For smooth metric measure spaces of vanishing $\Lambda_f(R, p)$ as $R \to \infty$, we show:

Theorem 6. *Let $\mathcal{F}$ be an orientable codimension-one foliation of $(M^{n+1}, \bar{g}, f)$ such that every orientable leaf L of $\mathcal{F}$ has the same constant f-mean curvature. If $\lim_{R\to\infty} = \Lambda_f(R, p) = 0$ for some $p \in M$, then leaves of $\mathcal{F}$ are f-minimal hypersurfaces of $(M^{n+1}, \bar{g}, f)$.*

We remark that the Gaussian space and $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$ enjoy the property that, for any point p, the ratio $\Lambda_f(R, p)$ vanishes as $R \to \infty$.

In Section 3, we prove:

Theorem 11. *Let $(M^{n+1}, \bar{g}, f)$ be an orientable smooth metric measure space and $\mathcal{F}$ a smooth codimension-one foliation of M by orientable leaves. If each leaf of $\mathcal{F}$ has the same constant f-mean curvature, then each leaf of $\mathcal{F}$ is strongly f-stable.*

Theorem 13. *There are no complete proper foliations in the Gaussian space $(\mathbb{R}^{n+1}, ds_0, f = |X|^2/2)$ whose leaves have the same constant f-mean curvature. In particular, there are no foliations of $\mathbb{R}^{n+1}$ whose leaves are complete proper self-similar solutions for mean curvature flow.*

2. Foliation whose leaves are f-minimal hypersurfaces

Let us start with the key lemma about the f-divergence of $\overline{\nabla}_\nu \nu$. The proof is analogous to that of Proposition 2.14 in [Barbosa et al. 1991], but we include its proof in the Appendix for the sake of completeness.

Lemma 1. *Let $\mathcal{F}$ be a smooth codimension-one foliation of a smooth metric measure space $(M^{n+1}, \bar{g}, f)$ and ν a unit normal vector field to the leaves of $\mathcal{F}$ in some open subset U of M. Define a tangent vector field $\xi = \overline{\nabla}_\nu \nu$. Then on U, we have:*

(a) $\overline{\mathrm{div}}_f \nu = -H_f$;

(b) $\overline{\mathrm{div}}_f \xi = \mathrm{div}_f\, \xi - |\xi|_g^2$;

(c) $\operatorname{div}_f \xi = |\xi|_g^2 + |A|^2 + \overline{\operatorname{Ric}}_f(\nu, \nu) - \nu H_f$.

Theorem 2. *Let $(M^{n+1}, \bar{g}, f)$ be a compact smooth metric measure space with nonnegative Bakry–Émery Ricci curvature and $\mathcal{F}$ a codimension-one smooth foliation of M whose leaves have the same constant f-mean curvature. Then every leaf of $\mathcal{F}$ is a totally geodesic and f-minimal hypersurface with vanishing Bakry–Émery Ricci curvature in the normal direction.*

Proof. Since H_f is constant in M, $\nu(H_f) \equiv 0$. Then Lemma 1(c) implies that

$$\operatorname{div}_f \xi = |A|^2 + |\xi|_g^2 + \overline{\operatorname{Ric}}_f(\nu, \nu)$$

on any leaf of $\mathcal{F}$, and therefore Lemma 1(b) implies that

$$\overline{\operatorname{div}}_f \xi = |A|^2 + \overline{\operatorname{Ric}}_f(\nu, \nu).$$

Recall that $dv_m = e^{-f}\, dv$. Integrating both sides and applying Stokes' theorem on M, we get

$$0 = \int_M \overline{\operatorname{div}}_f \xi \, dv_m = \int_M |A|^2 + \overline{\operatorname{Ric}}_f(\nu, \nu)\, dv_m,$$

that is, $|A|^2 = 0$ and $\overline{\operatorname{Ric}}_f(\nu, \nu) = 0$ on M. Therefore, every leaf is a totally geodesic hypersurface with vanishing Bakry–Émery Ricci curvature in the normal direction.

Since M is compact, there exists a point $m \in M$ such that $f(m) = \max_M f$. At m, we have $\overline{\nabla} f(m) = 0$. Therefore $H_f(L) = -\bar{g}(\overline{\nabla} f(m), \nu) = 0$, where L is the leaf which contains the point m. So, $H_f \equiv 0$ on any leaf of $\mathcal{F}$. This completes the proof. □

Remark 3. (1) The compactness condition in Theorem 2 is necessary. The smooth metric measure space $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$ has vanishing Bakry–Émery Ricci curvature and is noncompact. Translating solitons under the mean curvature flow do not change shape and are just translated in a direction with a constant speed. Up to rotating and scaling, they are represented by x_{n+1}-minimal hypersurfaces in the smooth metric measure space $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$ (see [Huisken and Sinestrari 1999]). By [Altschuler and Wu 1994] for $n = 2$, and [Gui, Jian and Ju 2010] for $n \geq 3$, there exists an entire rotationally symmetric strictly convex graphical hypersurface U, which gives a foliation by x_{n+1}-minimal hypersurfaces. But clearly U is not a totally geodesic hypersurface.

(2) The theorem of (Bonnet and) Myers [1941] says that a complete Riemannian manifold M is compact when M has Ricci curvature bounded from below by a positive constant. But this does not hold in general for a smooth metric measure space. One such example is the Gaussian space $(\mathbb{R}^{n+1}, ds_0, f = |X|^2/2)$. There are some generalizations of the Bonnet–Myers theorem with conditions on f [Morgan 2005; Wei and Wylie 2009].

Theorem 4. *Let $(M^{n+1}, \bar{g}, f)$ be a smooth metric measure space with positive Bakry–Émery Ricci curvature. Any smooth codimension-one foliation of M whose leaves have the same constant f-mean curvature cannot have a compact leaf.*

Proof. Suppose that, on the contrary, there exists a compact leaf L in the foliation $\mathcal{F}$. Lemma 1(c) implies that

$$\operatorname{div}_f \xi = |\xi|_g^2 + |A|^2 + \overline{\operatorname{Ric}}_f(\nu, \nu)$$

on L. Weighting both sides by $dA_m = e^{-f}\, dA$, integrating, and applying Stokes' theorem on L, we get a contradiction. □

Let $\mathcal{F}$ be a smooth orientable codimension-one foliation and L a leaf of $\mathcal{F}$. The weighted volume element $dA_m = \varphi_f$ of L is defined as follows:

$$\varphi_f(X_1, \dots, X_n) = e^{-f} g(X_1 \wedge \cdots \wedge X_n, \nu),$$

where the X_i are tangent vector fields ($i = 1, \dots, n$).

With a positively oriented frame field $\{e_1, \dots, e_n, e_{n+1} = \nu\}$, and its dual coframe $\{\omega_1, \dots, \omega_{n+1}\}$, the weighted volume elements $dA_m = \varphi_f$ and $dv_m = \Phi_m$ are expressed by

$$\varphi_f = e^{-f} \omega_1 \wedge \cdots \wedge \omega_n,$$
$$\Phi_f = e^{-f} \omega_1 \wedge \cdots \wedge \omega_{n+1}.$$

Both these weighted volume elements are related by the Rummler-type identity [Rummler 1979] as follows:

Lemma 5. *Let $(M^{n+1}, \bar{g}, f)$ be an orientable smooth metric measure space and $\mathcal{F}$ a smooth codimension-one foliation of M by orientable leaves. Then*

$$\bar{d}\varphi_f = (-1)^{n+1} H_f \Phi_f,$$

where φ_f is a weighted volume element of leaves of $\mathcal{F}$.

Proof. Taking exterior differentiation on φ_f, we have

$$\bar{d}\varphi_f = -e^{-f}\bar{d}f \wedge \omega_1 \wedge \cdots \wedge \omega_n + e^{-f}\bar{d}(\omega_1 \wedge \cdots \wedge \omega_n).$$

Since

$$\bar{d}f = e_1 f \omega_1 + \cdots + e_{n+1} f \omega_{n+1}$$

and

$$\bar{d}(\omega_1 \wedge \cdots \wedge \omega_n) = (-1)^{n+1} H \omega_1 \wedge \cdots \wedge \omega_{n+1},$$

we have

$$\begin{aligned}\bar{d}\varphi_f &= (-1)^{n+1} e^{-f}(e_{n+1} f)\omega_1 \wedge \cdots \wedge \omega_{n+1} + (-1)^{n+1} e^{-f} H \omega_1 \wedge \cdots \wedge \omega_{n+1} \\ &= (-1)^{n+1} H_f \Phi_f.\end{aligned}$$

□

Let p be a point in M, and $B_p(R)$ a geodesic ball in $(M, \bar{g})$ of radius R centered at p. The boundary of $B_p(R)$ is denoted by $\partial B_p(R)$. Define the ratio of the weighted volume of $B_p(R)$ and $\partial B_p(R)$ as follows:

$$\Lambda_f(R, p) = \frac{\mathrm{vol}_f(\partial B_p(R))}{\overline{\mathrm{vol}}_f(B_p(R))},$$

where $\overline{\mathrm{vol}}_f(B_p(R))$ and $\mathrm{vol}_f(\partial B_p(R))$ are the weighted volumes of $B_p(R)$ and $\partial B_p(R)$, respectively.

Theorem 6. *Let $\mathcal{F}$ be an orientable codimension-one foliation of $(M^{n+1}, \bar{g}, f)$ such that every orientable leaf L of $\mathcal{F}$ has the same constant f-mean curvature. If $\lim_{R\to\infty} \Lambda_f(R, p) = 0$ for some $p \in M$, then leaves of $\mathcal{F}$ are f-minimal hypersurfaces of $(M^{n+1}, \bar{g}, f)$.*

Proof. Suppose not. Then, choosing a normal vector field, we may assume that

$$(-1)^{n+1} H_f > 0.$$

Let σ_f be a weighted volume element of $\partial B_p(R)$. That is, for a local orthonormal frame field $\{X_1, \dots, X_n\}$ which is tangent to $\partial B_p(R)$,

$$\sigma_f(X_1, \dots, X_n) = e^{-f}.$$

On $\partial B_p(R)$, we have $\varphi_f \le \sigma_f$.

By Lemma 5, we have

$$\begin{aligned}\overline{\mathrm{vol}}_f(B_p(R)) &= \int_{B_p(R)} \Phi_f = \int_{B_p(R)} \frac{(-1)^{n+1}}{H_f} \bar{d}\varphi_f \\ &= \frac{(-1)^{n+1}}{H_f} \int_{\partial B_p(R)} \varphi_f \\ &\le \frac{(-1)^{n+1}}{H_f} \int_{\partial B_p(R)} \sigma_f \\ &= \frac{(-1)^{n+1}}{H_f} \mathrm{vol}_f(\partial B_p(R)).\end{aligned}$$

Therefore

$$0 < (-1)^{n+1} H_f \le \frac{\mathrm{vol}_f(\partial B_p(R))}{\overline{\mathrm{vol}}_f(B_p(R))} = \Lambda_f(R, p).$$

As R goes to ∞, we get a contradiction, and this completes the proof. □

Let $X = (x_1, \dots, x_{n+1})$ be the position vector in $\mathbb{R}^{n+1}$ and $|X|^2 = x_1^2 + \cdots + x_{n+1}^2$. Self-shrinkers under the mean curvature flow in $\mathbb{R}^{n+1}$ are represented by $|X|^2/2$-minimal hypersurfaces in the Gaussian space $(\mathbb{R}^{n+1}, ds_0, f = |X|^2/2)$ (see [Colding and Minicozzi 2012]).

By direct computation,

$$\lim_{R\to\infty} \frac{\operatorname{vol}_f(\partial B_p(R))}{\overline{\operatorname{vol}}_f(B_p(R))} = 0$$

in the Gaussian space, and therefore the following corollary is obtained:

Corollary 7. *Let $\mathcal{F}$ be an orientable codimension-one foliation of the Gaussian space such that every orientable leaf L of $\mathcal{F}$ has the same constant f-mean curvature. Then leaves of $\mathcal{F}$ are self-shrinkers.*

By direct computation,

$$\lim_{R\to\infty} \frac{\operatorname{vol}_f(\partial B_p(R))}{\overline{\operatorname{vol}}_f(B_p(R))} = 0$$

also holds in $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$, and therefore the following corollary is obtained:

Corollary 8. *Let $\mathcal{F}$ be an orientable codimension-one foliation of $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$ such that every orientable leaf L of $\mathcal{F}$ has the same constant f-mean curvature. Then leaves of $\mathcal{F}$ are translating solitons.*

Let Σ be a hypersurface in $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$ and H_f its f-mean curvature. Translating Σ in the direction of $e_{n+1} = (0, \dots, 0, 1) \in \mathbb{R}^{n+1}$, the f-mean curvature does not change. Using this property we get a Bernstein-type theorem for constant f-mean curvature surfaces.

Corollary 9. *Let $x_{n+1} = F(x_1, \dots, x_n)$ be a hypersurface of constant f-mean curvature defined on $\{x_{n+1} = 0\}$ in $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$. Then the f-mean curvature must be zero.*

Proof. Consider the graph graph(F) of the function F. The family

$$\{\operatorname{graph}(F) + te_{n+1}\}_{t\in\mathbb{R}}$$

gives a foliation whose leaves are hypersurfaces in $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$ with the same constant f-mean curvature. From Corollary 8, H_f vanishes. □

When f is a constant, then Corollary 9 becomes the corollary from p. 82 of [Chern 1965].

3. Stability of foliations whose leaves have the same constant f-mean curvature

Let Σ^n be a constant f-mean curvature hypersurface in $(M^{n+1}, \bar{g}, f)$. The f-stability operator L_f is defined as

$$L_f := \triangle_f + |A|^2 + \overline{\operatorname{Ric}}_f(\nu, \nu),$$

where ν is a unit normal vector field of Σ (see [Cheng et al. 2012; Colding and Minicozzi 2012; Espinar 2012]).

Definition 10. A two-sided hypersurface Σ in $(M^{n+1}, \bar{g}, f)$ with constant f-mean curvature is said to be strongly f-stable if for any compactly supported smooth function $u \in C_c^\infty(\Sigma)$, it satisfies

$$-\int_\Sigma uL_f u\, dA_m = \int_\Sigma |\nabla u|^2 - (|A|^2 + \overline{\mathrm{Ric}}_f(\nu, \nu))u^2\, dv_m \geq 0.$$

If Σ is an f-minimal hypersurface, then strong f-stability is equivalent to usual f-stability.

Theorem 11. *Let $(M^{n+1}, \bar{g}, f)$ be an orientable smooth metric measure space and $\mathscr{F}$ a smooth codimension-one foliation of M by orientable leaves. If each leaf of $\mathscr{F}$ has the same constant f-mean curvature, then each leaf of $\mathscr{F}$ is strongly f-stable.*

Proof. Let L be a leaf of $\mathscr{F}$ and u a smooth real-valued function which is compactly supported on a domain D in L (therefore, u is zero on ∂D). Then

$$-uL_f u = -u\triangle_f u - (|A|^2 + \overline{\mathrm{Ric}}_f(\nu, \nu))u^2 = -u\triangle_f u + u^2 \operatorname{div}_f \xi + u^2 |\xi|_g^2.$$

Here we apply equation (c) in Lemma 1.

Since $\operatorname{div}_f(u^2\xi) = 2ug(\nabla u, \xi) + u^2 \operatorname{div}_f \xi$, we get

$$-uL_f u = -u\triangle_f u - 2ug(\nabla u, \xi) + u^2|\xi|_g^2 + \operatorname{div}_f(u^2\xi).$$

Weighting both sides by dv_m, integrating over D, and applying Stokes' theorem twice for the first and the last terms, we have

$$\begin{aligned}\int_D -uL_f u\, dv_m &= \int_D -u\triangle_f u - 2ug(\nabla u, \xi) + u^2|\xi|_g^2 + \operatorname{div}_f(u^2\xi)\, dv_m \\ &= \int_D |\nabla u|_g^2 - 2ug(\nabla u, \xi) + u^2|\xi|_g^2\, dv_m \\ &= \int_D |\nabla u - u\xi|_g^2\, dv_m \geq 0.\end{aligned}$$

Since u is an arbitrary function, we conclude that L is f-stable. □

Remark 12. Let Σ be a graph over a domain $\Omega \subset \{x_{n+1} = 0\}$ in $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$ having constant f-mean curvature. Denote $\Sigma_t = \Sigma + te_{n+1}$, $t \in \mathbb{R}$. Then, by Theorem 11, every Σ_t is strongly f-stable. For example, the family of "grim reapers" $\Sigma_t = \{(x_1, \dots, x_n, t - \ln\cos x_1 : |x_1| < \pi/2)\}$ is a foliation in the open manifold $\{(x_1, \dots, x_{n+1}) : |x_1| < \pi/2)\}$ in $(\mathbb{R}^{n+1}, ds_0, f = x_{n+1})$. So, every grim reaper is strongly f-stable.

Let $\mathscr{F}$ be a foliation of the Gaussian space $(\mathbb{R}^{n+1}, ds_0, f = |X|^2/2)$. If every leaf of $\mathscr{F}$ is proper (respectively, complete), then $\mathscr{F}$ is said to be *proper* (respectively, *complete*).

Theorem 13. *There are no complete proper foliations in the Gaussian space* $(\mathbb{R}^{n+1}, ds_0, f = |X|^2/2)$ *whose leaves have the same constant f-mean curvature. In particular, there are no foliations of* $\mathbb{R}^{n+1}$ *whose leaves are complete proper self-similar solutions for mean curvature flow.*

Recall Colding and Minicozzi's result for self-shrinkers in the Gaussian space:

Theorem 14 [Colding and Minicozzi 2012]. *There are no f-stable complete self-shrinkers without boundary and with polynomial volume growth in the Gaussian space.*

Proof of Theorem 13. Suppose, on the contrary, that $\mathscr{F}$ is a complete, proper foliation whose leaves have the same f-mean curvature. By Corollary 7 and foliated structure, every leaf L of $\mathscr{F}$ is a self-shrinker without boundary. By Theorem 11, L is f-stable. Cheng and Zhou [2013] proved that for self-shrinkers, properness is equivalent to polynomial volume growth. Therefore, L is an f-stable complete self-shrinker without boundary and with polynomial volume growth in the Gaussian space. This contradicts Theorem 14. $\square$

Appendix

Let $\mathscr{F}$ be a smooth codimension-one foliation of a smooth metric measure space $(M^{n+1}, \bar{g}, f)$. On a leaf of $\mathscr{F}$, the induced metric is denoted by $g = i^*\bar{g}$, where i is the inclusion map. Let $\{e_1, \dots, e_{n+1}\}$ be a locally defined orthonormal frame field of the tangent bundle of M such that e_{n+1} is normal to the leaves of $\mathscr{F}$. Let us denote the dual coframe field by $\{\omega_1, \dots, \omega_{n+1}\}$, that is, $\omega_A(e_B) = \delta_{AB}$.

The connection one-forms ω_{AB} are given by exterior differentiation $\bar{d}$ of the ω_A, and are uniquely defined by Cartan's first structure equations:

$$\bar{d}\omega_A = \omega_{AB} \wedge \omega_B, \quad \omega_{AB} + \omega_{BA} = 0.$$

Cartan's second structure equations yield the curvature tensor

$$\bar{d}\omega_{AB} = \omega_{AC} \wedge \omega_{CB} + \Omega_{AB}, \tag{1}$$

where

$$\Omega_{AB} = \tfrac{1}{2} R_{ABCD}\omega_D \wedge \omega_C.$$

Throughout, we adopt Einstein's convention and the following indexing convention:

$$1 \le i, j, k, l \le n, \quad 1 \le A, B, C, D \le n+1.$$

The second fundamental form A of the leaves of $\mathscr{F}$ is given by

$$\omega_{n+1i} = -h_{ij}\omega_j, \tag{2}$$

where

$$h_{ij} = \bar{g}(A(e_i, e_j), e_{n+1}) = \bar{g}(\overline{\nabla}_{e_i} e_j, e_{n+1}).$$

The mean curvature is $H = \sum_i h_{ii}$.

Proof of Lemma 1. Consider an adapted orthonormal frame field $\{e_1, \dots, e_{n+1}\}$ on U such that $e_{n+1} = \nu$. We have

$$\begin{aligned} \overline{\mathrm{div}}_f \nu = e^f \,\overline{\mathrm{div}}(e^{-f}\nu) &= e^f \bar{g}(\overline{\nabla}_{e_A} e^{-f}\nu, e_A) \\ &= -\bar{g}(e_A f \nu, e_A) + \bar{g}(\overline{\nabla}_{e_A}\nu, e_A) \\ &= -\bar{g}(\overline{\nabla} f, \nu) - H = -H_f. \end{aligned}$$

Therefore, the equation (a) holds.

Furthermore,

$$\begin{aligned} \overline{\mathrm{div}}_f \xi &= e^f \,\overline{\mathrm{div}}(e^{-f}\xi) \\ &= e^f g(\overline{\nabla}_{e_i} e^{-f}\xi, e_i) + \bar{g}(\overline{\nabla}_{e_{n+1}} e^{-f}\xi, e_{n+1}) \\ &= \mathrm{div}_f\, \xi - \bar{g}(\xi, \overline{\nabla}_{e_{n+1}} e_{n+1}) = \mathrm{div}_f\, \xi - |\xi|_g^2. \end{aligned}$$

Therefore, the equation (b) holds.

Since $\bar{d}u = du + e_{n+1}(u)\omega_{n+1}$ for any smooth function u in U, from (2) we get

$$\omega_{n+1\,i} = -h_{ij}\omega_j + g(\xi, e_i)\omega_{n+1}. \tag{3}$$

On the one hand, from (1), we have

$$\begin{aligned} \bar{d}\omega_{n+1\,i} &= \omega_{n+1\,j} \wedge \omega_{ji} + R_{n+1\,i\,n+1\,k}\omega_k \wedge \omega_{n+1} + \tfrac{1}{2} R_{n+1ijk}\,\omega_k \wedge \omega_j \\ &= \big(-h_{jk}\omega_{ji}(e_{n+1}) - g(\xi, e_i)\omega_{ji}(e_k) + R_{n+1\,i\,n+1\,k}\big)\,\omega_k \wedge \omega_{n+1} \\ &\qquad + \text{terms with } \omega_k \wedge \omega_l. \end{aligned}$$

On the other hand, from (3),

$$\begin{aligned} \bar{d}\omega_{n+1i} &= -(dh_{ij} + e_{n+1}h_{ij}\omega_{n+1}) \wedge \omega_j - h_{ij}\omega_{jk} \wedge \omega_k - h_{ij}\omega_{jn+1} \wedge \omega_{n+1} \\ &\qquad + \bar{d}g(\xi, e_i)\,\omega_{n+1} + g(\xi, e_1)\,\omega_{n+1\,j} \wedge \omega_j \\ &= \big(e_{n+1}h_{ik} + h_{ij}\omega_{jk}(e_{n+1}) - h_{ij}h_{jk} + dg(\xi, e_i)(e_k) \\ &\qquad - g(\xi, e_i)g(\xi, e_j)\big)\,\omega_k \wedge \omega_{n+1} + \text{terms with } \omega_k \wedge \omega_l. \end{aligned}$$

By investigating both of the coefficients of $\omega_k \wedge \omega_{n+1}$ in $\bar{d}\omega_{n+1i}$, we have

$$g(\xi, e_i)g(\xi, e_k) + h_{ij}h_{jk} + R_{n+1in+1k} - (dh_{ik} + h_{ij}\omega_{jk} + h_{jk}\omega_{ji})(e_{n+1}) \\ = (\bar{d}g(\xi, e_i) + g(\xi, e_i)\omega_{ji})(e_k). \tag{4}$$

Since $\bar{d}g(\xi, e_i)(e_k) = dg(\xi, e_j)(e_k)$ and $g(\nabla_{e_i}\xi, e_i) = dg(\xi, e_i) + g(\xi, e_j)\omega_{ji}(e_i)$,

$$\begin{aligned}
\operatorname{div}_f \xi &= e^f \operatorname{div}(e^{-f}\xi) = \operatorname{div}\xi - g(\nabla f, \xi) \\
&= \sum_i (dg(\xi, e_i) + g(\xi, e_j)\omega_{ji})(e_i) - g(\nabla f, \xi) \\
&= \sum_i g(\xi, e_i)^2 + |A|^2 + \overline{\mathrm{Ric}}(\nu, \nu) - \nu H - g(\nabla f, \xi) \qquad \text{(by (4))} \\
&= |\xi|_g^2 + |A|^2 + \overline{\mathrm{Ric}}(\nu, \nu) - \nu H - \nu g(\overline{\nabla} f, \nu) + \nu g(\overline{\nabla} f, \nu) - g(\nabla f, \xi) \\
&= |\xi|_g^2 + |A|^2 - \nu H_f + \overline{\mathrm{Ric}}(\nu, \nu) + \overline{\nabla}^2 f(\nu, \nu) \\
&= |\xi|_g^2 + |A|^2 - \nu H_f + \overline{\mathrm{Ric}}_f(\nu, \nu).
\end{aligned}$$

This completes the proof. □

References

[Altschuler and Wu 1994] S. J. Altschuler and L. F. Wu, "Translating surfaces of the non-parametric mean curvature flow with prescribed contact angle", *Calc. Var. Partial Differential Equations* **2**:1 (1994), 101–111. MR 97b:58032 Zbl 0812.35063

[Bakry and Émery 1985] D. Bakry and M. Émery, "Diffusions hypercontractives", pp. 177–206 in *Séminaire de Probabilités, XIX* (Strasbourg, 1983–1984), edited by J. Azéma and M. Yor, Lecture Notes in Math. **1123**, Springer, Berlin, 1985. MR 88j:60131 Zbl 0561.60080

[Barbosa et al. 1987] J. L. M. Barbosa, J. M. Gomes, and A. M. Silveira, "Foliation of 3-dimensional space forms by surfaces with constant mean curvature", *Bol. Soc. Brasil. Mat.* **18**:2 (1987), 1–12. MR 90j:53054 Zbl 0747.53029

[Barbosa et al. 1991] J. L. M. Barbosa, K. Kenmotsu, and G. Oshikiri, "Foliations by hypersurfaces with constant mean curvature", *Math. Z.* **207**:1 (1991), 97–107. MR 92b:53034 Zbl 0731.53033

[Cheng and Zhou 2013] X. Cheng and D. Zhou, "Volume estimate about shrinkers", *Proc. Amer. Math. Soc.* **141**:2 (2013), 687–696. MR 2996973 Zbl 1262.53030

[Cheng et al. 2012] X. Cheng, T. Mejia, and D. Zhou, "Eigenvalue estimate and compactness for closed f-minimal surfaces", preprint, 2012. arXiv 1210.8448

[Chern 1965] S.-S. Chern, "On the curvatures of a piece of hypersurface in Euclidean space", *Abh. Math. Sem. Univ. Hamburg* **29** (1965), 77–91. MR 32 #6376 Zbl 0147.20901

[Colding and Minicozzi 2012] T. H. Colding and W. P. Minicozzi, II, "Generic mean curvature flow, I: Generic singularities", *Ann. of Math.* (2) **175**:2 (2012), 755–833. MR 2993752 Zbl 1239.53084

[Espinar 2012] J. M. Espinar, "Manifolds with density, applications and gradient Schrödinger operators", preprint, 2012. arXiv 1209.6162

[Gui, Jian and Ju 2010] C. Gui, H. Jian, and H. Ju, "Properties of translating solutions to mean curvature flow", *Discrete Contin. Dyn. Syst.* **28**:2 (2010), 441–453. MR 2011h:35081 Zbl 1193.35085

[Huisken and Sinestrari 1999] G. Huisken and C. Sinestrari, "Convexity estimates for mean curvature flow and singularities of mean convex surfaces", *Acta Math.* **183**:1 (1999), 45–70. MR 2001c:53094 Zbl 0992.53051

[Lawson 1974] H. B. Lawson, Jr., “Foliations”, *Bull. Amer. Math. Soc.* **80** (1974), 369–418. MR 49 #8031 Zbl 0293.57014

[Meeks 1988] W. H. Meeks, III, “The topology and geometry of embedded surfaces of constant mean curvature”, *J. Differential Geom.* **27**:3 (1988), 539–552. MR 89h:53025 Zbl 0617.53007

[Morgan 2005] F. Morgan, “Manifolds with density”, *Notices Amer. Math. Soc.* **52**:8 (2005), 853–858. MR 2006g:53044 Zbl 1118.53022

[Myers 1941] S. B. Myers, “Riemannian manifolds with positive mean curvature”, *Duke Math. J.* **8** (1941), 401–404. MR 3,18f Zbl 0025.22704

[Oshikiri 1981] G.-I. Oshikiri, “A remark on minimal foliations”, *Tôhoku Math. J.* (2) **33**:1 (1981), 133–137. MR 83b:57017 Zbl 0437.57013

[Pyo 2014] J. Pyo, “Compact translating solitons with non-empty planar boundary”, preprint, 2014.

[Rummler 1979] H. Rummler, “Quelques notions simples en géométrie Riemannienne et leurs applications aux feuilletages compacts”, *Comment. Math. Helv.* **54**:2 (1979), 224–239. MR 80m:57021 Zbl 0409.57026

[Wei and Wylie 2009] G. Wei and W. Wylie, “Comparison geometry for the Bakry–Emery Ricci tensor”, *J. Differential Geom.* **83**:2 (2009), 377–405. MR 2011a:53064 Zbl 1189.53036

Received May 16, 2013. Revised April 8, 2014.

JUNCHEOL PYO
DEPARTMENT OF MATHEMATICS
PUSAN NATIONAL UNIVERSITY
BUSAN 609-735
SOUTH KOREA

jcpyo@pusan.ac.kr

and

SCHOOL OF MATHEMATICS
KOREA INSTITUTE FOR ADVANCED STUDY (KIAS)
SEOUL 130-722
SOUTH KOREA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 271, No. 1, 2014
dx.doi.org/10.2140/pjm.2014.271.243

ON THE EXISTENCE OF LARGE DEGREE GALOIS REPRESENTATIONS FOR FIELDS OF SMALL DISCRIMINANT

JEREMY ROUSE AND FRANK THORNE

Let L/K be a Galois extension of number fields. We prove two lower bounds on the maximum of the degrees of the irreducible complex representations of $\mathrm{Gal}(L/K)$, the sharper of which is conditional on the Artin conjecture and the generalized Riemann hypothesis. Our bound is nontrivial when $[K : \mathbb{Q}]$ is small and L has small root discriminant, and might be summarized as saying that such fields can't be "too abelian".

1. Introduction

It is known that the discriminant of a number field cannot be too small. Minkowski's work on the geometry of numbers implies that

$$|\mathrm{Disc}(K)| > \left(\frac{e^2\pi}{4} - o(1)\right)^{[K:\mathbb{Q}]};$$

we write this bound as $\mathrm{rd}_K > e^2\pi/4 - o(1)$, where $\mathrm{rd}_K := (|\mathrm{Disc}(K)|)^{1/[K:\mathbb{Q}]}$ is the *root discriminant* of K. These bounds can be improved by using analytic properties of the Dedekind zeta function of K, and this was noticed by Stark (see the parenthetical comment in the proof of Lemma 4 on page 140 of [Stark 1974]), and worked out in detail by Andrew Odlyzko [1976] in his MIT dissertation (supervised by Stark). The sharpest known bounds, due to Poitou [1977] (see also [Odlyzko 1990]), are

$$\mathrm{rd}_K \geq (60.8395\ldots)^{r_1/[K:\mathbb{Q}]}(22.3816\ldots)^{2r_2/[K:\mathbb{Q}]} - O([K:\mathbb{Q}]^{-2/3}), \tag{1-1}$$

where $[K : \mathbb{Q}] = r_1 + 2r_2$, and r_1 and r_2 are the numbers of real and complex embeddings of K, respectively. (The error term in (1-1) can be improved.) If one assumes the generalized Riemann hypothesis, the constants above can be improved to $215.3325\ldots$ and $44.7632\ldots$ respectively.

Conversely, Golod and Shafarevich [1964] proved that these bounds are sharp apart from the constants, by establishing the existence of *infinite class field towers*

MSC2010: primary 11R29; secondary 11R42.
Keywords: Artin L-function, Galois representation, Rankin–Selberg L-function.

$K_1 \subseteq K_2 \subseteq K_3 \subseteq \cdots$ where each K_{i+1}/K_i is abelian and unramified, so that each field K_i has the same root discriminant. Martinet [1979] gave the example $K_1 = \mathbb{Q}(\zeta_{11} + \zeta_{11}^{-1}, \sqrt{-46})$, which has an infinite 2-class field tower of root discriminant $92.2\ldots$, and Hajir and Maire [2001; 2002] constructed a tower of fields with root discriminants bounded by 82.2, in which tame ramification is allowed.

It is expected that fields of small discriminant should be uncommon. For example, Odlyzko [1990] asked whether there are infinitely many fields of *prime degree* of bounded root discriminant; such fields cannot be constructed via class field towers. Several researchers have studied this question in small degree. Jones and Roberts [2007] studied the set of Galois number fields $K/\mathbb{Q}$ with certain fixed Galois groups G; for a variety of groups, including $A_4, A_5, A_6, S_4, S_5, S_6$, they proved that $\mathrm{rd}_K > 44.7632\ldots$ apart from a finite list of fields K which they compute explicitly. Voight [2008] studied the set of all totally real number fields K with $\mathrm{rd}_K \leq 14$, finding that there are exactly 1229 such fields, each with $[K:\mathbb{Q}] \leq 9$.

In light of this work, it is natural to ask whether Galois extensions of small absolute discriminant must have any special algebraic properties. (The analogous problems for nonnormal extensions are much more delicate.) The easiest result to prove is that they *cannot be abelian*, and we carry this out over $\mathbb{Q}$ in the introduction (starting with (1-4)). In [Leshin 2013], it is proven that, given a number field K, a positive integer n, and real number N, there are only finitely many Galois extensions L/K with $\mathrm{Gal}(L/K)$ solvable with derived length $\leq n$ and with $\mathrm{rd}_L \leq N$. In this paper, we study the representation theory of Galois groups of extensions of small discriminant, and prove that such Galois groups must have (relatively) large degree complex representations.

We will prove two versions of this result. The first is the following:

Theorem 1.1. *Let L/K be a Galois extension and let r be the maximum of the degrees of the irreducible complex representations of* $\mathrm{Gal}(L/K)$. *Then there is a constant C_1 so that*

$$r \geq \frac{1}{\log \mathrm{rd}_L}\left(C_1 \frac{\log\log[L:\mathbb{Q}]}{\log\log\log[L:\mathbb{Q}]} - \log[K:\mathbb{Q}]\right). \tag{1-2}$$

Remark. The bound of course only makes sense for large $[L:\mathbb{Q}]$. A straightforward but somewhat lengthy calculation shows that we may take $C_1 = \frac{1}{16}$ provided $[L:\mathbb{Q}] \geq e^{e^8}$.

The basic idea of the proof is to regard L as an abelian extension of an intermediate field F of small degree. The existence of such an F follows from Theorem 12.23 of [Isaacs 2006], which states that if G is a finite group with the property that all of its irreducible representations have degree $\leq r$, then G must have an abelian subgroup of index $\leq (r!)^2$. (There is also a converse given in Problem 2.9 of [Isaacs 2006]: if G has an irreducible representation of degree $> r$, then G cannot have an abelian subgroup of index $\leq r$.) We may then adapt our proof for the abelian case to prove that either L has small root discriminant, or F has relatively large degree.

It is also possible to study the representations of $\mathrm{Gal}(L/\mathbb{Q})$ directly, without first passing to an intermediate extension F, via Artin L-functions. We were unable to improve upon Theorem 1.1 this way, but under the hypothesis that Artin L-functions are well behaved we prove the following improvement of Theorem 1.1:

Theorem 1.2. *Assume that all Artin L-functions are entire and satisfy the Riemann hypothesis. There is a positive constant C_2 so that if L/K is any Galois extension of number fields of degree d, then* $\mathrm{Gal}(L/K)$ *must have an irreducible complex representation of degree at least*

$$\frac{C_2(\log[L:\mathbb{Q}])^{1/5}}{(\log \mathrm{rd}_L)^{2/5}[K:\mathbb{Q}]^{3/5}}. \tag{1-3}$$

Remark. Two issues arise when attempting to prove an unconditional version of this result. The first is that the unconditional zero-free regions for L-functions have implied constants that depend quite badly on the degree of the L-function involved. (See for example Theorem 5.33 of [Iwaniec and Kowalski 2004].) The second is the presence of the possible exceptional zero. Without accounting for the exceptional zero issue, it seems that the best lower bound we can obtain using the zero-free regions mentioned above is $r \gg \sqrt{\log\log[L:\mathbb{Q}]}$, for an implied constant depending on K and on rd_L.

We now illustrate the nature of our question by handling the case where $L/\mathbb{Q}$ is abelian of degree > 2. By Kronecker–Weber we have $L \subseteq \mathbb{Q}(\zeta_n)$ for some n, and

$$\zeta_{L/\mathbb{Q}}(s) = \zeta(s) \prod_{i=2}^{[L:\mathbb{Q}]} L(s, \chi_i), \tag{1-4}$$

where χ_i are Dirichlet characters of conductor N_i for some $N_i | n$. We have $\mathrm{Disc}(L) = \prod_i N_i$, and therefore

$$\log \mathrm{rd}_L = \frac{1}{[L:\mathbb{Q}]} \sum_{i=2}^{[L:\mathbb{Q}]} \log N_i. \tag{1-5}$$

Let $M := \sqrt{[L:\mathbb{Q}]/2}$. There are at most M^2 Dirichlet characters with conductor $\leq M$, so that with $[L:\mathbb{Q}] = 2M^2$ the right side of (1-5) is greater than $\frac{1}{2}\log M$, so that

$$\mathrm{rd}_L > \exp\left(\tfrac{1}{4}\log\left(\tfrac{1}{2}[L:\mathbb{Q}]\right)\right), \tag{1-6}$$

a bound of the same shape as our theorems. Although this proof is not complicated, it makes essential use of class field theory and it seems that the use of sophisticated tools cannot be avoided. We could improve our bound somewhat, but note that it is already stronger than the (conditional) bound $\log \mathrm{rd}_L \geq C_2(\log[L:\mathbb{Q}])^{1/5}$ implied by Theorem 1.2. Observe also that for $L = \mathbb{Q}(\zeta_p)$ we have $\mathrm{rd}_L = p^{(p-2)/(p-1)}$ and $[L:\mathbb{Q}] = p-1 \approx \mathrm{rd}_L$, implying a limit on the scope for improvement.

As an application of Theorem 1.2 we can say something about unramified extensions of a fixed number field K. Of course, the maximal unramified *abelian* extension of K is the Hilbert class field of K and the degree of this extension is h_K, the order of the ideal class group. However, there are number fields K with Galois extensions L unramified at all finite primes so that $\mathrm{Gal}(L/K)$ has no nontrivial abelian quotients. One of Artin's favorite examples is $K = \mathbb{Q}(\sqrt{2869})$, where if L is the splitting field of $x^5 - x - 1$ over $\mathbb{Q}$, then L/K is unramified and $\mathrm{Gal}(L/K) \cong A_5$.

Corollary 1.3. *Assume that all Artin L-functions are entire and satisfy the Riemann hypothesis. Let $L_1/K, L_2/K, \dots, L_N/K$ be linearly disjoint unramified Galois extensions and suppose that* $\mathrm{Gal}(L_i/K)$ *has an irreducible representation of degree r for* $1 \leq i \leq N$. *Then there is a constant C_3 so that*

$$\log N \leq C_3 r^5 \log^2(|\mathrm{Disc}(K)|)[K:\mathbb{Q}].$$

Remark. The main theorem proven in [Ellenberg and Venkatesh 2006] shows that the number M of degree n unramified extensions of K satisfies

$$\log M \ll_\epsilon n^\epsilon \big(n \log|\mathrm{Disc}(K)| + C_4[K:\mathbb{Q}]\big)$$

for a constant C_4 depending on n. Because the power of $\log|\mathrm{Disc}(K)|$ is smaller, this result is better for fixed n and varying K. However, since the size of C_4 is not specified, our result is better for fixed K and varying n.

Remark. Another potential application occurs in the case when $r = 2$ and $K = \mathbb{Q}$. Our theorem gives bounds on the number of degree-2 Artin L-functions with conductor bounded by q. In the odd case, these arise from weight-1 newforms of level q, and in the even case, these arise (conjecturally) from Maass forms with eigenvalue $\frac{1}{4}$. However, we obtain bounds that are worse than polynomial in q. Michel and Venkatesh [2002] used the Petersson–Kuznetsov formula to obtain bounds of the form $q^{c+\epsilon}$, where c is a constant depending on the type of representation (dihedral, tetrahedral, octahedral, or icosahedral).

Throughout the paper we use the notation $|\mathrm{Disc}(K)|$ for the absolute value of the discriminant of K, $\mathbb{O}_K$ for the ring of integers of K, $\mathrm{rd}_K = |\mathrm{Disc}(K)|^{1/[K:\mathbb{Q}]}$, $\mathcal{N}_{K/\mathbb{Q}}(\mathfrak{a})$ for the norm from K to $\mathbb{Q}$ of an ideal of $\mathbb{O}_K$, h_K for the order of the ideal class group of $\mathbb{O}_K$, and $\mathfrak{d}_{L/K}$ for the relative discriminant of L over K. We denote by $C_1, C_2, \dots$ a sequence of absolute constants. We also occasionally write $f \ll g$ to mean $f \leq Cg$ for some constant C, absolute unless otherwise noted.

We provide a little bit of preliminary background in Section 2, and then we prove Theorem 1.1 in Section 3 and Theorem 1.2 and Corollary 1.3 in Section 4.

2. Background on number fields, discriminants, and conductors

In this section we briefly recall a few facts related to zeta and L-functions associated to number fields, used in the proofs of both Theorem 1.1 and Theorem 1.2.

The *Dedekind zeta function* of a number field L is given by the Dirichlet series

$$\zeta_L(s) = \sum \mathcal{N}_{L/\mathbb{Q}}(\mathfrak{a})^{-s}, \tag{2-1}$$

where the sum is over integral ideals of $\mathcal{O}_L$. For a Galois extension L/K, this zeta function enjoys the factorization

$$\zeta_L(s) = \zeta_{L/K}(s) = \prod_{\rho\in\mathrm{Irr}(\mathrm{Gal}(L/K))} L(s,\rho)^{\deg\rho}, \tag{2-2}$$

where ρ varies over all irreducible complex representations of $G := \mathrm{Gal}(L/K)$, and $L(s,\rho)$ is the associated Artin L-function. (For background on Artin L-functions see [Neukirch 1999]; see p. 524 for the proof of (2-2) in particular.)

This formula is the nonabelian generalization of (1-4). In general, it is not known that the $L(s,\rho)$ are "proper" L-functions (as defined on [Iwaniec and Kowalski 2004, p. 94] for example) and in particular that they are holomorphic in the critical strip. However, this was conjectured by Artin; we refer to this assumption as the Artin conjecture and assume its truth in Section 4.

Remark. As a consequence of Brauer's theorem on group characters [Neukirch 1999, p. 522], it is known that the Artin L-functions are quotients of Hecke L-functions, and therefore meromorphic, and this suffices in many applications. For example, Lagarias and Odlyzko [1977] used this fact to prove an unconditional and effective version of the Chebotarev density theorem.

If $\mathrm{Gal}(L/K)$ is abelian, then the representations are all one-dimensional, and class field theory establishes that the characters of $\mathrm{Gal}(L/K)$ coincide with Hecke characters of L/K, so that (2-2) becomes

$$\zeta_{L/K}(s) = \zeta_K(s) \prod_{i=2}^{[L:K]} L(s,\chi_i), \tag{2-3}$$

where the product ranges over Hecke characters of K. As in our application of (1-4), we will argue that there cannot be too many characters χ or representations ρ of small conductor (and, in the latter case, of bounded degree).

We can use (2-2) to derive more general versions of (1-5): it follows [Neukirch 1999, p. 527] from (2-2) that the relative discriminant $\mathfrak{d}(L/K)$ satisfies the formula

$$\mathfrak{d}(L/K) = \prod_{\rho \in \mathrm{Irr}(G)} \mathfrak{f}(\rho)^{\deg \rho}, \tag{2-4}$$

where the ideal $\mathfrak{f}(\rho)$ of K is the Artin conductor associated to ρ.

If L/K is abelian then we can write this as $\mathfrak{d}(L/K) = \prod_{\chi_i} \mathfrak{f}(\chi_i)$. Taking norms down to $\mathbb{Q}$ and using the relation [Neukirch 1999, p. 202]

$$|\mathrm{Disc}(L)| = |\mathrm{Disc}(K)|^{[L:K]} \mathcal{N}_{K/\mathbb{Q}}(\mathfrak{d}_{L/K}), \tag{2-5}$$

we obtain

$$\log \mathrm{rd}_L = \log \mathrm{rd}_K + \frac{1}{[L:\mathbb{Q}]} \sum_i \log \mathcal{N}_{K/\mathbb{Q}} \mathfrak{f}(\chi_i). \tag{2-6}$$

If L/K is not necessarily abelian, then the conductor $q(\rho)$ of $L(s, \rho)$ is related to $\mathfrak{f}(\rho)$ by the formula

$$q(\rho) = |\mathrm{Disc}(K)|^{\deg \rho} \mathcal{N}_{K/\mathbb{Q}} \mathfrak{f}(\rho). \tag{2-7}$$

Taking absolute norms in (2-4), multiplying by $|\mathrm{Disc}(K)|^d$, and again using (2-5) we obtain

$$|\mathrm{Disc}(L)| = \prod_{\rho \in \mathrm{Irr}(G)} q(\rho)^{\deg \rho}. \tag{2-8}$$

3. Proof of Theorem 1.1

We first prove a lemma bounding some quantities which occur in the proof.

Lemma 3.1. *For a number field F of degree f, the following hold*:

(1) *The number of ideals $\mathfrak{a}$ of $\mathbb{O}_F$ with $\mathcal{N}(\mathfrak{a}) < Y$ is bounded by $eY(1 + \log Y)^f$.*

(2) *We have $h_F < e|\mathrm{Disc}(F)|^{1/2}\bigl(1 + \frac{1}{2}\log|\mathrm{Disc}(F)|\bigr)^f$.*

Proof. This is standard and we give an easy proof inspired by [Cojocaru and Murty 2006, p. 68]. We have that $\zeta_F(s) = \sum_{n=1}^{\infty} a_n(F)/n^s$, where $a_n(F)$ is the number of integral ideals of norm n in $\mathbb{O}_F$. The coefficientwise bound $\zeta_F(s) < \zeta(s)^f = \sum_n d_f(n) n^{-s}$ yields that, for $\sigma > 1$,

$$\sum_{n<Y} d_f(n) < \sum_n d_f(n)(Y/n)^{\sigma} = Y^{\sigma} \zeta(\sigma)^f.$$

We now choose $\sigma = 1 + 1/\log Y$, and use the fact that $\zeta(\sigma) < 1 + \frac{1}{\sigma - 1}$ for $\sigma > 1$.

The second part follows from the classical Minkowski bound (see for example [Neukirch 1999, Chapter 1.6]), which implies that each ideal class in $\mathbb{O}_F$ is represented by an ideal $\mathfrak{a}$ with $\mathcal{N}(\mathfrak{a}) < \sqrt{|\mathrm{Disc}(F)|}$. □

Proof of Theorem 1.1. The proof is similar to that of (1-6), but we will need to work with messier inequalities.

By the character theory remarks after the theorem, L has a subfield F for which L/F is abelian, such that $[F : K] \leq (r!)^2 < r^{2r}$. We assume that $[L : \mathbb{Q}]$ and therefore rd_L are bounded below by absolute constants ($[L : \mathbb{Q}] \geq e^{e^8}$ suffices). Depending on the relative sizes of these quantities, we will see that either

$$r \geq C_5 \frac{\log\log[L : \mathbb{Q}]}{\log\log\log[L : \mathbb{Q}]} - \log[K : \mathbb{Q}] \tag{3-1}$$

or

$$\log \mathrm{rd}_L \geq C_6 \log\log[L : \mathbb{Q}] \tag{3-2}$$

for positive constants C_5 and C_6, implying the theorem. There is no obstacle to determining particular values for these constants, but for simplicity we omit the details.

We begin with the generalization (2-6) of (1-5), which said that

$$\log \mathrm{rd}_L = \log \mathrm{rd}_F + \frac{1}{[L : \mathbb{Q}]} \sum_i \log \mathcal{N}_{F/\mathbb{Q}} \mathfrak{f}(\chi_i), \tag{3-3}$$

where χ_i are distinct Hecke characters of F. The number of characters of conductor $\mathfrak{m}$ is less than $2^{[F:\mathbb{Q}]} h_F \mathcal{N}(\mathfrak{m})$ [Milne 2011, Theorem V.1.7, p. 146], and Lemma 3.1 bounds both h_F and the number of $\mathfrak{m}$ which can appear, so that for $Y \geq 1$ the number of characters whose conductor has norm $\leq Y$ is bounded above by $e^2 Y^2 |\mathrm{Disc}(F)|^{1/2} \big(2 + \log(Y^2 |\mathrm{Disc}(F)|)\big)^{2[F:\mathbb{Q}]}$.

Given $[L : \mathbb{Q}]$ and $[F : \mathbb{Q}]$, suppose that $Y > e^{-1} |\mathrm{Disc}(F)|^{-1/2}$ is defined by the equation

$$\frac{[L : \mathbb{Q}]}{2[F : \mathbb{Q}]} = e^2 Y^2 |\mathrm{Disc}(F)|^{1/2} \big(2 + \log(Y^2 |\mathrm{Disc}(F)|)\big)^{2[F:\mathbb{Q}]}, \tag{3-4}$$

so that in (3-3) there are at least $\dfrac{[L : \mathbb{Q}]}{2[F : \mathbb{Q}]}$ characters of conductor $> Y$, and hence

$$\log \mathrm{rd}_L \geq \log \mathrm{rd}_F + \frac{1}{2[F : \mathbb{Q}]} \log Y. \tag{3-5}$$

(Observe that we do not necessarily have $Y > 1$, for example if L is the Hilbert class field of F.) We divide our analysis of (3-5) into three cases and prove that each implies (3-1) or (3-2).

Large discriminant. If $\mathrm{Disc}(F) \geq [L : \mathbb{Q}]^{1/10}$, we ignore (3-5) and instead note that $\log \mathrm{rd}_F \geq \log[L : \mathbb{Q}]/(10[F : \mathbb{Q}])$, and so

$$\log \mathrm{rd}_L \geq \frac{1}{10 r^{2r} [K : \mathbb{Q}]} \log[L : \mathbb{Q}], \tag{3-6}$$

and we obtain at least one of (3-1) and (3-2) depending on whether $r^{2r}[K:\mathbb{Q}] > (\log[L:\mathbb{Q}])^{1/2}$ or not.

We assume henceforth that $\mathrm{Disc}(F) < [L:\mathbb{Q}]^{1/10}$, which implies that $[F:\mathbb{Q}] < \frac{1}{10}\log[L:\mathbb{Q}]$ for $[F:\mathbb{Q}] \geq 3$, and write $Y' := \max(Y, 100)$ and $Z := Y'^2\,|\mathrm{Disc}(F)|^{1/2}$.

Small discriminant and large degree. Assume to start with that either $Z \leq (2 + 2\log Z)^{2[F:\mathbb{Q}]}$ or $Y < 100$. Applying our upper bounds on $\mathrm{Disc}(F)$, $[F:\mathbb{Q}]$, and Z, we see that

$$[L:\mathbb{Q}]^{4/5} \leq C_7(4\log(Y'^2\,|\mathrm{Disc}(F)|))^{4[F:\mathbb{Q}]}. \tag{3-7}$$

Taking logarithms and applying the bound[1] $Y'^2\,|\mathrm{Disc}(F)| < [L:\mathbb{Q}]$, we obtain

$$\log[L:\mathbb{Q}] \leq C_8[F:\mathbb{Q}]\log\log[L:\mathbb{Q}], \tag{3-8}$$

so that $\log[L:\mathbb{Q}]/\log\log[L:\mathbb{Q}] \leq C_9[K:\mathbb{Q}]r^{2r}$, which implies that

$$r \geq C_{10}\frac{\log\log[L:\mathbb{Q}]}{\log\log\log[L:\mathbb{Q}]} - \log[K:\mathbb{Q}].$$

Small discriminant and small degree. Finally, assume that $Z > (2+2\log Z)^{2[F:\mathbb{Q}]}$ and $Y \geq 100$. Then $[F:\mathbb{Q}] \leq C_{11}\log Z/\log\log Z$, and our bound on $\mathrm{Disc}(F)$ implies that $\log Z \leq C_{12}\log Y$, so that $[F:\mathbb{Q}] \leq C_{13}\log Y/\log\log Y$. We thus have $\log Y/(2[F:\mathbb{Q}]) \geq C_{14}\log\log Y$, and so, by (3-5),

$$\log \mathrm{rd}_L \geq \log \mathrm{rd}_F + \frac{1}{2[F:\mathbb{Q}]}\log Y \geq \log \mathrm{rd}_F + C_{14}\log\log Y.$$

Finally, (3-4) implies that $\log Y \leq \frac{1}{2}\log[L:\mathbb{Q}]$, giving us

$$\log \mathrm{rd}_L \geq \log \mathrm{rd}_F + C_{15}\log\log[L:\mathbb{Q}]. \tag{3-9}$$

This completes our list of cases, and hence the proof. □

4. Proof of Theorem 1.2

In the proof we will assume familiarity with Artin L-functions and Rankin–Selberg convolutions, as described in [Neukirch 1999] (and Section 2) and [Iwaniec and Kowalski 2004], respectively. We also assume the truth of the Artin conjecture. There is no theoretical obstacle to carrying out the methods of this section without any unproved hypotheses, but when we tried this the error terms in (4-6) were too large to be of interest.

As in the nonabelian case, we need to bound the number of possible $q(\rho)$ of bounded conductor (and now also of bounded degree). However, in general the representations ρ are not (yet!) known to correspond to arithmetic objects which

[1] If $Y' = 100$, this follows from $\mathrm{Disc}(F) < [L:\mathbb{Q}]^{1/10}$. If $Y' = Y$, this follows from (3-4).

might be more easily counted. Instead, in Proposition 4.1 we (conditionally) bound the number of possible L-functions. Assuming GRH and the Artin conjecture, we will see that any two such Artin L-functions must have rather different Dirichlet series representations, because their Rankin–Selberg convolution cannot have a pole. A pigeonhole-type argument will then allow us to bound the number of possible L-functions.

After proving Proposition 4.1 we will conclude as before. In brief, if L/K is a Galois extension of large degree with many representations of small degree, then many of these representations will have large conductor, and so L will have large discriminant.

Proposition 4.1. *Assume that the Artin conjecture and Riemann hypothesis hold for Artin L-functions, and let ρ and ρ' be distinct irreducible nontrivial representations of* $\mathrm{Gal}(L/K)$ *of degree r and conductor $\leq q$ (as defined in* (2-7)*). The Artin L-series of ρ and ρ' have Euler products*

$$L(s,\rho)=\prod_{\mathfrak{p}}\prod_{i=1}^{d}\bigl(1-\alpha_{i,\rho}(\mathfrak{p})N(\mathfrak{p})^{-s}\bigr)^{-1},\quad L(s,\rho')=\prod_{\mathfrak{p}}\prod_{i=1}^{d}\bigl(1-\alpha_{i,\rho'}(\mathfrak{p})N(\mathfrak{p})^{-s}\bigr)^{-1}.$$

Assume that $\log q > r[K:\mathbb{Q}]$. *Then, for $X \geq C_{16} r^2 \log^2 q$, we have*

$$\sum_{\substack{\mathcal{N}\mathfrak{p}\in[X,2X]\\ \mathfrak{p}\text{ unramified}}}\ \sum_{1\leq i\leq r}|\alpha_{i,\rho}(\mathfrak{p})-\alpha_{i,\rho'}(\mathfrak{p})| \geq \frac{X}{2r\log X}. \tag{4-1}$$

Furthermore, the number of representations of degree $\leq r$ and conductor $\leq q$ is at most

$$C_{17}^{r^3\log^2(q)[K:\mathbb{Q}]}, \tag{4-2}$$

for an absolute constant C_{17}.

Remark. The bound (4-2) is rather simple-minded, and we could remove the factor $[K:\mathbb{Q}]$ by instead insisting that q be sufficiently large in terms of K.

Proof. This is essentially Proposition 5.22 of [Iwaniec and Kowalski 2004]; although our conclusion is stronger, our proof is essentially the same.

Consider the tensor product representations $\rho\otimes\bar{\rho}$ and $\rho\otimes\overline{\rho'}$, whose L-functions are equal to the Rankin–Selberg convolutions $L(s,\rho\otimes\bar{\rho})$ and $L(s,\rho\otimes\overline{\rho'})$ (so that the notation is not ambiguous). A simple character-theoretic argument shows that the trivial representation does not occur in $\rho\otimes\overline{\rho'}$, while it occurs with multiplicity one in $\rho\otimes\bar{\rho}$. Assuming the Artin conjecture, then, $L(s,\rho\otimes\overline{\rho'})$ and $L(s,\rho\otimes\bar{\rho})\zeta(s)^{-1}$ are entire functions.

Let ϕ be a smooth test function with support in $[1,2]$, image in $[0,1]$, and $\widehat{\phi}(0)=\int_1^2\phi(t)\,dt\in\bigl(\frac{3}{4},1\bigr)$; throughout this section, all implied constants (including

C_{16}, etc.) depend on our fixed choice of ϕ. Also, let $X \geq 2$, with stricter lower bounds to be imposed later. Then, using the "explicit formula" [ibid., Theorem 5.11] we find (see p. 118 of [ibid.]) that, assuming GRH and the Artin conjecture,

$$\left|\sum_n \Lambda_{\rho\otimes\overline{\rho'}}(n)\phi(n/X)\right| \ll \sqrt{X}\log \mathfrak{q}(\rho\otimes\overline{\rho'}), \tag{4-3}$$

$$\left|\sum_n \Lambda_{\rho\otimes\bar{\rho}}(n)\phi(n/X) - \widehat{\phi}(0)X\right| \ll \sqrt{X}\log \mathfrak{q}(\rho\otimes\bar{\rho}), \tag{4-4}$$

where the coefficients Λ are defined, for any L-function $L(f,s)$, by the relation

$$\sum_n \Lambda_f(n)n^{-s} = -\frac{L'}{L}(f,s).$$

Also, $\mathfrak{q}(\rho) = q(\rho)\prod_{j=1}^r(|\kappa_j|+3)$ denotes the analytic conductor of ρ defined by equation (5.8) of [ibid.]. To bound this analytic conductor, we require information about the gamma factors of Artin L-functions and the conductor of $\rho\otimes\psi$ (where $\psi = \bar{\rho}$ or $\overline{\rho'}$).

The fact that the $L(s,\rho)$ are factors of the Dedekind zeta function and the fact that the Dedekind zeta function only has gamma factors $\Gamma(s/2)$ and $\Gamma((s+1)/2)$ imply that $0 \leq \kappa_j \leq 1/2$ for all j. We have $q(\rho\otimes\psi) = |\mathrm{Disc}(K)|^{\deg(\rho\otimes\psi)}\mathcal{N}_{K/\mathbb{Q}}(\mathfrak{f}(\rho\otimes\psi))$, where

$$\mathfrak{f}(\rho\otimes\psi) = \prod_{\mathfrak{p}\nmid\infty}\mathfrak{p}^{f_{\mathfrak{p}}(\rho\otimes\psi)}, \qquad f_{\mathfrak{p}}(\rho\otimes\psi) = \sum_{i=0}^{\infty}\frac{g_i}{g_0}\mathrm{codim}\ V_{\rho\otimes\psi}^{G_i}.$$

Here $V_{\rho\otimes\psi}$ is a vector space affording the representation $\rho\otimes\psi$, G_i is the i-th ramification group, and $g_i = |G_i|$ (this definition is from p. 527 of [Neukirch 1999]). It is easy to see that $\mathrm{codim}\ V_{\rho\otimes\psi}^{G_i} \leq r\ \mathrm{codim}\ V_{\rho}^{G_i} + r\ \mathrm{codim}\ V_{\psi}^{G_i}$. It follows from this and the formulas

$$f_{\mathfrak{p}}(\rho) = \sum_{i=0}^{\infty}\frac{g_i}{g_0}\mathrm{codim}\ V_{\rho}^{G_i} \quad\text{and}\quad f_{\mathfrak{p}}(\psi) = \sum_{i=0}^{\infty}\frac{g_i}{g_0}\mathrm{codim}\ V_{\psi}^{G_i}.$$

that $f_{\mathfrak{p}}(\rho\otimes\psi) \leq r(f_{\mathfrak{p}}(\rho)+f_{\mathfrak{p}}(\psi))$ and

$$q(\rho\otimes\psi) \leq |\mathrm{Disc}(K)|^{r^2}\mathcal{N}_{K/\mathbb{Q}}(\mathfrak{f}(\rho)\mathfrak{f}(\psi)). \tag{4-5}$$

Combining these estimates for the analytic conductor yields the bound

$$\log \mathfrak{q}(\rho\otimes\psi) \leq \bigl(2r^2 + r^2\log|\mathrm{Disc}(K)| + r\log\mathcal{N}_{K/\mathbb{Q}}(\mathfrak{f}(\rho)\mathfrak{f}(\psi))\bigr) \leq 3r\log q.$$

Let $\alpha_{i,\rho}$ and $\alpha_{i,\rho'}$ be the Frobenius eigenvalues for ρ and ρ' respectively, for $1 \leq i \leq r$. Then, unpacking the definition of the Rankin–Selberg convolution (or, equivalently, the tensor product representation), we conclude that

$$\text{(4-6)}\quad \sum_{\mathcal{N}\mathfrak{p}\in[X,2X]}\ \sum_{1\le i,j\le r} \left|\alpha_{i,\rho}(\mathfrak{p})\overline{\alpha_{j,\rho}(\mathfrak{p})} - \alpha_{i,\rho}(\mathfrak{p})\overline{\alpha_{j,\rho'}(\mathfrak{p})}\right| \ge \frac{\widehat{\phi}(0)X}{\log(2X)} - \frac{C_{18}\sqrt{X}}{\log X}(r\log q).$$

The sum above is over prime ideals; we have removed the prime powers which contribute (assuming X is large) less than

$$\frac{2r^2[K:\mathbb{Q}]\sqrt{X}}{\log X} < \frac{2r\log(q)\sqrt{X}}{\log X},$$

which is contained in the error term above. Next, we remove the terms coming from "ramified primes", those for which $\alpha_{i,\rho}(\mathfrak{p}) = 0$ for some i. These occur precisely at the primes for which $\mathfrak{p}\,|\,\mathfrak{f}(\rho)$, and the number of these between X and $2X$ is at most $\log q/\log X$. Noting that for a prime $\mathfrak{p}$ that is unramified in L/K, $|\alpha_{i,\rho}(\mathfrak{p})| = 1$ for all i, we get

$$\text{(4-7)}\quad \sum_{\substack{\mathcal{N}\mathfrak{p}\in[X,2X]\\ \mathfrak{p}\text{ unramified}}}\ \sum_{1\le j\le r} |\alpha_{j,\rho}(\mathfrak{p}) - \alpha_{j,\rho'}(\mathfrak{p})| \ge \frac{\widehat{\phi}(0)X}{r\log(2X)} - C_{19}\left(\frac{\sqrt{X}\log q + r\log q}{\log X}\right).$$

If $X \ge C_{16}r^2\log^2(q)$ with $C_{16} = \max(2^{14}, 100C_{19}^2)$, the error term above is at most $X/(5r\log X)$ and the main term is at least $7X/(10r\log X)$, establishing the first part of the proposition.

The second part follows easily from the first. Let M be the number of primes in the above sum; then, if

$$|\alpha_{i,\rho}(\mathfrak{p}) - \alpha_{i,\rho'}(\mathfrak{p})| \le \frac{X}{2r^2\log(X)M}$$

for all i and unramified $\mathfrak{p}$, then (4-1) is contradicted. Note that since $\mathfrak{p}$ is unramified, $\alpha_{i,\rho}(\mathfrak{p})$ and $\alpha_{i,\rho'}(\mathfrak{p})$ lie on the unit circle. No more than $2\pi Y$ points can be placed on the unit circle with pairwise distances at least $1/Y$. Hence, by the pigeonhole principle, there can be at most

$$N = \left(\frac{4\pi r^2\log(X)M}{X}\right)^{r(\pi_K(2X)-\pi_K(X))}$$

Artin L-functions of degree r and conductor $\le q$. Here, $\pi_K(X)$ is the number of prime ideals of $\mathbb{O}_K$ of norm at most X. We have $M \le \pi_K(2X) - \pi_K(X) \le 2[K:\mathbb{Q}]X/\log X$ and so

$$\log N \le \frac{2rX[K:\mathbb{Q}]}{\log X}(\log(8\pi r^2) + \log[K:\mathbb{Q}]).$$

We have

$$\log X \geq \log 2^{14} + 2\log r + 2\log\log q > \log 8\pi + 2\log r + \log[K:\mathbb{Q}],$$

and hence

$$\log N \leq 2C_{16} r^3 \log^2(q)[K:\mathbb{Q}].$$

As $e^{r^3 \log^2(q)[K:\mathbb{Q}]}$ increases rapidly in r, the number of representations of conductor at most q and degree $r' \leq r$ is bounded by $C_{17}^{r^3 \log^2(q)[K:\mathbb{Q}]}$, for some absolute constant C_{17}. □

We now prove Theorem 1.2 using Proposition 4.1.

Proof of Theorem 1.2. Assume L/K is a Galois extension whose irreducible complex representations all have degree at most r. Choose the smallest $A \geq e$ such that

$$\frac{[L:K]}{2r^2} \leq \sum_{i=1}^{r} C_{17}^{r^2(r-i)\log^2(A^2)[K:\mathbb{Q}]}. \tag{4-8}$$

We wish to estimate A in terms of $|\mathrm{Disc}(L)|$. By Proposition 4.1, the number of representations with degree $r-i$ and conductor $\leq A^{2r/(r-i)}$ is at most

$$C_{17}^{(r-i)^3 \log^2(A^{2r/(r-i)})[K:\mathbb{Q}]} = C_{17}^{4r^2(r-i)\log^2(A^2)[K:\mathbb{Q}]}.$$

Every other representation has $q(\rho)^{\deg\rho} \geq A^{2r}$. There are at least $[L:K]/2r^2$ of these, and so (2-8) gives

$$|\mathrm{Disc}(L)| = \prod_{\rho\in\mathrm{Irr}(\mathrm{Gal}(L/K))} q(\rho)^{\deg\rho} \geq (A^{2r})^{[L:K]/2r^2} = A^{[L:K]/r}.$$

Thus, $\log A \leq (r/[L:K])\log|\mathrm{Disc}(L)| = r\cdot[K:\mathbb{Q}]\log \mathrm{rd}_L$.

Equation (4-8) gives

$$\frac{[L:K]}{2r^2} \leq 2C_{17}^{r^3\log^2(A^2)[K:\mathbb{Q}]},$$

enlarging C_{17} if necessary so that $C_{17}^{e^2} > 2$, which gives

$$\begin{aligned}\log[L:K] &\leq \log(4r^2) + r^3\log^2(A)[K:\mathbb{Q}]\log C_{17}\\ &\leq (\log 4 + \log C_{17})\cdot r^5[K:\mathbb{Q}]^3(\log \mathrm{rd}_L)^2.\end{aligned}$$

Hence, there is an absolute constant C_2 so that

$$r \geq C_2 \frac{(\log[L:K])^{1/5}}{[K:\mathbb{Q}]^{3/5}(\log \mathrm{rd}_L)^{2/5}}.$$ □

Finally, we prove Corollary 1.3 using Proposition 4.1.

Proof of Corollary 1.3. Let L be the compositum of the L_i. Then we have

$$\mathrm{Gal}(L/K) = \prod_{i=1}^{N} \mathrm{Gal}(L_i/K).$$

From Theorem 4.21 of [Isaacs 2006], if $\rho_i : \mathrm{Gal}(L_i/K) \to \mathrm{GL}_r(\mathbb{C})$ is an irreducible representation, then the map $\tilde{\rho}_i(g) = \rho_i(g|_{L_i})$ is also an irreducible representation of $\mathrm{Gal}(L/K)$ which is distinct from $\tilde{\rho}_j$ for $i \neq j$. All of these representations have conductor $q = |\mathrm{Disc}(K)|^r$, and Proposition 4.1 implies that there is an absolute constant C_3 so that

$$\log N \leq C_3 r^5 \log^2(|\mathrm{Disc}(K)|)\,[K : \mathbb{Q}],$$

as desired. □

Acknowledgments

We would like to thank Kannan Soundararajan for suggesting the basic idea of this paper to us. We'd also like to thank Charlotte Euvrard and the referee for a number of very helpful suggestions.

References

[Cojocaru and Murty 2006] A. C. Cojocaru and M. R. Murty, *An introduction to sieve methods and their applications*, London Mathematical Society Student Texts **66**, Cambridge University Press, 2006. MR 2006k:11184 Zbl 1121.11063

[Ellenberg and Venkatesh 2006] J. S. Ellenberg and A. Venkatesh, "The number of extensions of a number field with fixed degree and bounded discriminant", *Ann. of Math.* (2) **163**:2 (2006), 723–741. MR 2006j:11159 Zbl 1099.11068

[Golod and Shafarevich 1964] E. S. Golod and I. R. Shafarevich, "О башне полей классов", *Izv. Akad. Nauk SSSR Ser. Mat.* **28**:2 (1964), 261–272. Translated as "On the class field tower" in *Amer. Math. Soc. Transl.* **48** (1965), 91–102. MR 28 #5056 Zbl 0136.02602

[Hajir and Maire 2001] F. Hajir and C. Maire, "Tamely ramified towers and discriminant bounds for number fields", *Compositio Math.* **128**:1 (2001), 35–53. MR 2002g:11149 Zbl 1042.11072

[Hajir and Maire 2002] F. Hajir and C. Maire, "Tamely ramified towers and discriminant bounds for number fields, II", *J. Symbolic Comput.* **33**:4 (2002), 415–423. MR 2003h:11137 Zbl 1086.11051

[Isaacs 2006] I. M. Isaacs, *Character theory of finite groups*, AMS Chelsea, Providence, RI, 2006. MR 2270898 Zbl 1119.20005

[Iwaniec and Kowalski 2004] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications **53**, American Mathematical Society, Providence, RI, 2004. MR 2005h:11005 Zbl 1059.11001

[Jones and Roberts 2007] J. W. Jones and D. P. Roberts, "Galois number fields with small root discriminant", *J. Number Theory* **122**:2 (2007), 379–407. MR 2008e:11140 Zbl 1163.11079

[Lagarias and Odlyzko 1977] J. C. Lagarias and A. M. Odlyzko, "Effective versions of the Chebotarev density theorem", pp. 409–464 in *Algebraic number fields: L-functions and Galois properties* (Durham, 1975), edited by A. Fröhlich, Academic Press, London, 1977. MR 56 #5506 Zbl 0362.12011

[Leshin 2013] J. Leshin, "Solvable number field extensions of bounded root discriminant", *Proc. Amer. Math. Soc.* **141**:10 (2013), 3341–3352. MR 3080157 Zbl 06203295

[Martinet 1979] J. Martinet, "Petits discriminants", *Ann. Inst. Fourier* (*Grenoble*) **29**:1 (1979), 159–170. MR 81h:12006 Zbl 0387.12006

[Michel and Venkatesh 2002] P. Michel and A. Venkatesh, "On the dimension of the space of cusp forms associated to 2-dimensional complex Galois representations", *Int. Math. Res. Not.* **2002**:38 (2002), 2021–2027. MR 2003i:11064 Zbl 1008.11019

[Milne 2011] J. S. Milne, "Class field theory", course notes, v. 4.01, 2011, http://www.jmilne.org/math/CourseNotes/cft.html.

[Neukirch 1999] J. Neukirch, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften **322**, Springer, Berlin, 1999. MR 2000m:11104 Zbl 0956.11021

[Odlyzko 1976] A. M. Odlyzko, "Lower bounds for discriminants of number fields", *Acta Arith.* **29**:3 (1976), 275–297. MR 53 #5531 Zbl 0286.12006

[Odlyzko 1990] A. M. Odlyzko, "Bounds for discriminants and related estimates for class numbers, regulators and zeros of zeta functions: a survey of recent results", *Sém. Théor. Nombres Bordeaux* (2) **2**:1 (1990), 119–141. MR 91i:11154 Zbl 0722.11054

[Poitou 1977] G. Poitou, "Minorations de discriminants (d'après A. M. Odlyzko)", pp. 136–153 (Exp. No. 479) in *Séminaire Bourbaki, Vol. 1975/76*, Lecture Notes in Math. **567**, Springer, Berlin, 1977. MR 55 #7995 Zbl 0359.12010

[Stark 1974] H. M. Stark, "Some effective cases of the Brauer–Siegel theorem", *Invent. Math.* **23** (1974), 135–152. MR 49 #7218 Zbl 0278.12005

[Voight 2008] J. Voight, "Enumeration of totally real number fields of bounded root discriminant", pp. 268–281 in *Algorithmic number theory* (Banff, 2008), edited by A. J. van der Poorten and A. Stein, Lecture Notes in Comput. Sci. **5011**, Springer, Berlin, 2008. MR 2010a:11228 Zbl 1205.11125

Received March 4, 2013. Revised June 17, 2014.

JEREMY ROUSE
DEPARTMENT OF MATHEMATICS
WAKE FOREST UNIVERSITY
P.O. BOX 7388
WINSTON-SALEM, NC 27109
UNITED STATES

rouseja@wfu.edu

FRANK THORNE
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTH CAROLINA
1523 GREENE STREET
COLUMBIA, SC 29208
UNITED STATES

thorne@math.sc.edu

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

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 will 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, which should be used sparingly.

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.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 271 No. 1 September 2014

0030-8730(201409)271:1;1-1