Pacific Journal of Mathematics

Volume 308 No. 1 September 2020

PACIFIC JOURNAL OF MATHEMATICS

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

msp.org/pjm

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

The subscription price for 2020 is US $520/year for the electronic version, and $705/year for print and electronic.
Subscriptions, requests for back issues and changes of subscriber 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 1945-5844 electronic, 0030-8730 printed) 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. 308, No. 1, 2020
https://doi.org/10.2140/pjm.2020.308.1

ON THE TOPOLOGICAL DIMENSION OF THE GROMOV BOUNDARIES OF SOME HYPERBOLIC Out(F_N)-GRAPHS

MLADEN BESTVINA, CAMILLE HORBEZ AND RICHARD D. WADE

We give upper bounds, linear in the rank, to the topological dimensions of the Gromov boundaries of the intersection graph, the free factor graph and the cyclic splitting graph of a finitely generated free group.

Introduction

The curve graph $\mathcal{C}(\Sigma)$ of an orientable hyperbolic surface of finite type Σ is an essential tool in the study of the mapping class group of Σ. It was proved to be Gromov hyperbolic by Masur and Minsky [1999], and its Gromov boundary was identified with the space of ending laminations on Σ by Klarreich [1999]. A striking application of the curve graph and its geometry at infinity is the recent proof by Bestvina, Bromberg and Fujiwara [Bestvina et al. 2015] that Mod(Σ) has finite asymptotic dimension, which implies in turn that it satisfies the integral Novikov conjecture. A crucial ingredient in the proof is the finite asymptotic dimension of the curve graph: this was first proved by Bell and Fujiwara [2008] using Masur and Minsky's *tight geodesics*, and recently recovered (with a bound linear in the genus of Σ and the number of punctures) by Bestvina and Bromberg [2019], via finite capacity dimension of the Gromov boundary. The latter approach builds upon work

MSC2010: 20F28, 20F65.
Keywords: free factor graph, free factor complex, boundary, topological dimension, automorphism, free group.

of Gabai [2009], who bounded the topological dimension of $\partial_\infty \mathcal{C}(\Sigma)$ by building covers in terms of train-tracks on the surface; bounding the capacity dimension requires getting more metric control on such covers.

The importance of the curve graph in the study of mapping class groups led people to look for hyperbolic graphs with $\mathrm{Out}(F_N)$-actions. In the present paper, we will be mainly interested in three of them: the *free factor graph* FF_N, the *cyclic splitting graph* FZ_N and the *intersection graph* I_N. The Gromov boundary of each graph is homeomorphic to a quotient of a subspace of the boundary ∂CV_N of outer space [Bestvina and Reynolds 2015; Hamenstädt 2012; Horbez 2016a; Dowdall and Taylor 2017]. As ∂CV_N has dimension equal to $3N-5$ [Bestvina and Feighn 1992; Gaboriau and Levitt 1995] and the quotient maps are cell-like, this bounds the cohomological dimension of each of these Gromov boundaries. A priori this does not imply finiteness of their topological dimensions. This is the goal of the present paper.

Main Theorem. *Let* $N \geq 2$*.*

- *The boundary* $\partial_\infty I_N$ *has topological dimension at most* $2N-3$*.*
- *The boundary* $\partial_\infty FF_N$ *has topological dimension at most* $2N-2$*.*
- *The boundary* $\partial_\infty FZ_N$ *has topological dimension at most* $3N-5$*.*

We do not know whether equality holds in any of these cases (see the open questions at the end of the introduction). Following Gabai's work, our proof relies on constructing a decomposition of each Gromov boundary in terms of a notion of *train-tracks*. We hope that, by getting further control on the covers we construct, this approach may pave the way towards a proof of finite asymptotic dimension of FF_N, FZ_N or I_N (via finite capacity dimension of their Gromov boundaries).

The rest of the introduction is devoted to explaining the strategy of our proof. Although we treat the cases of $\partial_\infty I_N$, $\partial_\infty FF_N$ and $\partial_\infty FZ_N$ all at once in the paper, we will mainly focus on the free factor graph in this introduction for simplicity, and only say a word about how the proof works for FZ_N (the proof for I_N is similar).

Bounding the topological dimension. To establish our main theorem, we will use the following two topological facts [Engelking 1978, Lemma 1.5.2 and Proposition 1.5.3]. Let X be a separable metric space.

(1) If X can be written as a finite union $X = X_0 \cup X_1 \cup \cdots \cup X_k$ where each X_i is 0-dimensional, then $\dim(X) \leq k$.

Zero-dimensionality of each X_i will be proved by appealing to the following fact.

(2) If there exists a countable cover of X_i by closed 0-dimensional subsets, then $\dim(X_i) = 0$.

For example, one can recover the fact that $\dim(\mathbb{R}^n) \leq n$ from the above two facts, by using the decomposition of $\mathbb{R}^n$ where X_i is the set of points having exactly i

rational coordinates. Zero-dimensionality of each X_i can be proved using the second point, by decomposing each X_i into countably many closed subsets X_i^j, two points being in the same set precisely when they have the same rational coordinates (one easily checks that these sets X_i^j are 0-dimensional by finding arbitrarily small boxes around each point of X_i^j with boundary outside of X_i^j).

Notice that a decomposition (hereafter, a *stratification*) of X into finitely many subsets X_i as in the first point can be provided by a map $X \to \{0, \dots, k\}$ (hereafter, an *index map*). Showing that each X_i is zero-dimensional amounts to proving that every point in X_i has clopen neighborhoods (within X_i) of arbitrary small diameter (equivalently, every point in X_i has arbitrary small open neighborhoods in X_i with empty boundary). Actually, thanks to the second point above, it is enough to write each stratum X_i as a countable union of closed subsets (called hereafter a *cell decomposition*) and prove 0-dimensionality of each cell of the decomposition.

Stratification and cell decomposition of $\partial_\infty FF_N$. The boundary $\partial_\infty FF_N$ is a separable metric space when equipped with a visual metric. Points in the boundary are represented by F_N-trees, and a first reasonable attempt could be to define a stratification of $\partial_\infty FF_N$ by using an index map similar to the one introduced in [Gaboriau and Levitt 1995], that roughly counts orbits of branch points and of directions at these points in the trees. Although we make use of these features, our definition of the stratification is slightly different; it is based on a notion of train-tracks.

A train-track τ (see Definition 2.1) consists of a triple $(S^\tau, \sim_V, \sim_D)$, where S^τ is an F_N-action on a simplicial tree, and $\sim_V$ (resp. $\sim_D$) is an equivalence relation on the set of vertices (resp. directions) in S^τ. We say that a tree $T \in \overline{cv}_N$ is *carried* by τ if there is an F_N-equivariant map $f : S^\tau \to T$ (called a *carrying map*) that is in a sense compatible with the train-track structure. The typical situation is when f identifies two vertices in S^τ if and only if they are $\sim_V$-equivalent, and identifies the germs of two directions based at equivalent vertices in S^τ if and only if the directions are $\sim_D$-equivalent. For technical reasons, our general definition of a carrying map is slightly weaker; in particular it is a bit more flexible about possible images of vertices in S^τ at which there are only three equivalence classes of directions. A point $\xi \in \partial_\infty FF_N$ — which is an equivalence class of trees that all admit alignment-preserving bijections to one another — is then *carried* by τ if some (equivalently, any) representative is carried. If τ carries a point in $\partial_\infty FF_N$ then its underlying tree S^τ has a free F_N-action and determines an open simplex in the interior of outer space.

A train-track τ as above has an index $i(\tau)$ of at most $2N-2$: this is a combinatorial datum which mainly counts orbits of equivalence classes of vertices and directions in τ. The *index* $i(\xi)$ of a point $\xi \in \partial_\infty FF_N$ is then defined as the maximal index of a train-track that carries ξ. A train-track determines a *cell* $P(\tau)$ in $\partial_\infty FF_N$, defined as the set of points $\xi \in \partial_\infty FF_N$ such that τ carries ξ and $i(\tau) = i(\xi)$.

We define a stratification of $\partial_\infty FF_N$ by letting X_i be the collection of points of index i. Each X_i is covered by the countable collection of the cells $P(\tau)$, where τ varies over all train-tracks of index i. In view of the topological facts recalled above, our main theorem follows for $\partial_\infty FF_N$ from the following points:

(1) (Proposition 3.1) The boundary of $P(\tau)$ in $\partial_\infty FF_N$ is contained in a union of cells of strictly greater index. This implies that $P(\tau)$ is closed in $X_{i(\tau)}$.

(2) (Proposition 4.1) Each cell $P(\tau)$ has dimension at most 0.

We say that $P(\tau)$ has dimension "at most" 0, and not "equal to" 0, as we do not exclude the possibility that $P(\tau)$ is empty.

For the first point, we show that if $(\xi_n)_{n\in\mathbb{N}} \in P(\tau)^{\mathbb{N}}$ converges to $\xi \in \partial P(\tau)$, then the carrying maps f_n from S^τ to representatives of ξ_n converge to a map f from S^τ to a representative T of ξ. However, the limiting map f may no longer be a carrying map: for example, inequivalent vertices in S^τ may have the same image in the limit, and edges in S^τ may be collapsed to a point by f. We then collapse all edges in S^τ that map to a point under f and get an induced map f' from the collapse S' to T. This determines a new train-track structure τ' on S'. A combinatorial argument enables us to count the number of directions "lost" when passing from τ to τ', and show that $i(\tau') > i(\tau)$ unless T is carried by τ.

Our proof of the second point relies on a *cell decomposition* process similar to Gabai's [2009], who used *splitting sequences* of train-tracks on surfaces to get finer and finer covers of the space of ending laminations. In our context, starting from a cell $P(\tau)$, we will construct finer and finer decompositions of $P(\tau)$ into clopen subsets by means of *folding sequences* of train-tracks. Eventually the subsets in the decomposition have small enough diameter.

More precisely, starting from a train-track τ one can "resolve" an illegal turn in τ by folding it. There are several possibilities for the folded track (see the figures in Section 4), so this operation yields a subdivision of $P(\tau)$ into various $P(\tau^j)$. A first crucial fact we prove is that these $P(\tau^j)$ are all open in $P(\tau)$. Now, if $\xi \in P(\tau)$ and $\varepsilon > 0$, then by folding the train-track τ for sufficiently long we can reach a train-track τ' with $\mathrm{diam}(P(\tau')) < \varepsilon$. Here, hyperbolicity of FF_N is crucial: our folding sequence will determine an (unparametrized) quasigeodesic going to infinity (towards ξ) in FF_N. The set $P(\tau_i)$ defined by the train-track obtained at time i of the process is contained in the set of endpoints of geodesic rays in FF_N starting at the simplicial tree associated to τ_0 and passing at a bounded distance from τ_i. From the definition of the visual metric on the boundary, this implies that the diameter of $P(\tau_i)$ converges to 0 as we move along the folding path. All in all, we find a train-track τ' such that $P(\tau')$ is an open neighborhood of ξ in $P(\tau)$, and $\mathrm{diam}(P(\tau')) < \varepsilon$. The boundary of $P(\tau')$ in $P(\tau)$ is empty (it contains points of strictly greater index), so we get $\dim(P(\tau)) = 0$, as required.

A word on $\partial_\infty FZ_N$. Some points in $\partial_\infty FZ_N$ are represented by trees in which some free factor system of F_N is elliptic. This leads us to work with train-tracks that are allowed to have bigger vertex stabilizers, and the index of a train-track now also takes into account the complexity of this elliptic free factor system. A priori, our definition of index in this setting only yields a quadratic bound in N to the topological dimension of $\partial_\infty FZ_N$, but we then get the linear bound from cohomological dimension. We also have to deal with the fact that a point $\xi \in \partial_\infty FZ_N$ is an equivalence class of trees that may not admit alignment-preserving bijections between different representatives; this leads us to analyzing preferred (mixing) representatives of these classes more closely.

Organization of the paper. In Section 1, we review the definitions of FF_N, I_N and FZ_N, and the descriptions of their Gromov boundaries. We also prove a few facts concerning mixing representatives of points in $\partial_\infty FZ_N$, which are used to tackle the last difficulty mentioned in the above paragraph. Train-tracks and indices are then defined in Section 2, and the stratifications of the Gromov boundaries are given. We prove in Section 3 that each cell $P(\tau)$ is closed in its stratum, by showing that $\partial P(\tau)$ is made of points with strictly higher index. We introduce folding moves in Section 4, and use them to prove that each cell $P(\tau)$ has dimension at most 0. The proof of our main theorem is then completed in Section 5. In the Appendix we illustrate the motivations behind some technical requirements that appear in our definitions of train-tracks and carrying maps.

Some open questions. As mentioned earlier, we hope that the cell decompositions of the boundaries defined in the present paper will provide a tool to tackle the question of finiteness of the asymptotic dimension of the various graphs, following the blueprint of [Bestvina and Bromberg 2019].

Gabai [2009] used the cell decomposition of the ending lamination space $\mathcal{EL}$ to show that it is highly connected for sufficiently complicated surfaces (connectivity had been established earlier for most surfaces by Leininger and Schleimer [2009]). It is unknown whether the boundaries of I_N, FF_N or FZ_N are connected. The question of local connectivity of each boundary (or the related question of local connectivity of the boundary of the Culler–Vogtmann outer space) is also open to our knowledge.

We would also like to address the question of finding lower bounds on the dimensions. Gabai showed that the topological dimension of $\mathcal{EL}$ is bounded from below by $3g + p - 4$ (where g is the genus of the surface and p is the number of punctures). Since the ending lamination space of a once-punctured surface sits as a subspace of $\partial_\infty FF_{2g}$ and $\partial_\infty FZ_{2g}$, this gives a lower bound on the dimension of $\partial_\infty FF_{2g}$ and $\partial_\infty FZ_{2g}$. Improving the gap between the upper and lower bounds, as well as finding a lower bound for $\partial_\infty I_N$, is an interesting problem.

The free splitting complex has been notably absent from the above discussion. This is because its boundary is harder to describe, and does not appear as a quotient of a subspace of the boundary of outer space. However, we expect that our methods can be applied here when its boundary is better understood. One can also apply these techniques to the boundaries of relative versions of the various $\mathrm{Out}(F_N)$-complexes [Horbez 2016a; Guirardel and Horbez 2019].

Recent developments. After the first version of this paper was completed, Guirardel and the first two named authors proved a related result about the asymptotic geometry of $\mathrm{Out}(F_N)$, namely that $\mathrm{Out}(F_N)$ is boundary amenable [Bestvina et al. 2017]. In particular, we now know that $\mathrm{Out}(F_N)$ satisfies the Novikov conjecture. From the surface viewpoint, Hamenstädt [2019] has shown that asymptotic dimension of the disk graph of a handlebody of genus at least two is at most quadratic in the genus. Also, Bestvina, Bromberg and Fujiwara [Bestvina et al. 2019] gave a proof that the mapping class group has finite asymptotic dimension that does not directly appeal to the finite asymptotic dimension of the curve complex (although curve complexes and the Masur–Minsky distance formula still play vital roles). Finiteness of the asymptotic dimension of $\mathrm{Out}(F_N)$ and its related complexes is still a wide open problem.

1. Hyperbolic Out(F_N)-graphs and their boundaries

We review the definitions of three hyperbolic $\mathrm{Out}(F_N)$-graphs (the free factor graph, the intersection graph and the $\mathcal{Z}$-splitting graph) and the descriptions of their Gromov boundaries. The only novelties are some facts concerning mixing $\mathcal{Z}$-averse trees in the final subsection.

1A. ***The free factor graph.*** The *free factor graph* FF_N is the graph (equipped with the simplicial metric) whose vertices are the conjugacy classes of proper free factors of F_N. Two conjugacy classes $[A]$ and $[B]$ of free factors are joined by an edge if they have representatives A, B such that $A \subsetneq B$ or $B \subsetneq A$. Its hyperbolicity was proved in [Bestvina and Feighn 2014]. When $N = 2$ each proper free factor is the conjugacy class of a primitive element. In order to make this graph connected, one adds an edge in FF_2 between two conjugacy classes if they have representatives that form a basis of F_2.

To describe its Gromov boundary, we first recall that *unprojectivized outer space* cv_N is the space of all F_N-equivariant isometry classes of minimal, free, simplicial, isometric F_N-actions on simplicial metric trees. Its closure $\overline{cv}_N$ (for the equivariant Gromov–Hausdorff topology) was identified in [Cohen and Lustig 1995; Bestvina and Feighn 1992] with the space of all minimal *very small* F_N-trees, i.e., F_N-actions on $\mathbb{R}$-trees in which all stabilizers of nondegenerate arcs are cyclic (possibly trivial) and root-closed, and tripod stabilizers are trivial. There exists a coarsely Out(F_N)-equivariant map $\pi : cv_N \to FF_N$, which sends a tree T to a free factor that is elliptic in a tree $\overline{T}$ obtained from T by collapsing some edges to points.

An F_N-tree $T \in \partial cv_N$ is *arational* if no proper free factor of F_N is elliptic in T, or acts with dense orbits on its minimal subtree. Two arational trees T and T' are *equivalent* ($T \sim T'$) if there exists an F_N-equivariant alignment-preserving bijection from T to T'. We denote by $\mathcal{AT}_N$ the subspace of ∂cv_N made of arational trees. Arational trees were introduced by Reynolds [2012], who proved that every arational F_N-tree is either free, or dual to an arational measured lamination on a once-holed surface.

Theorem 1.1 [Bestvina and Reynolds 2015; Hamenstädt 2012]. *There is a homeomorphism*

$$\partial\pi : \mathcal{AT}_N/\!\sim\; \to \partial_\infty FF_N,$$

which extends the map π continuously to the boundary. By this we mean that for all sequences $(S_n)_{n\in\mathbb{N}} \in (cv_N)^{\mathbb{N}}$ converging to a tree $S_\infty \in \mathcal{AT}_N$ (for the topology on $\overline{cv}_N$), the sequence $(\pi(S_n))_{n\in\mathbb{N}}$ converges to $\partial\pi(S_\infty)$ (for the topology on $FF_N \cup \partial_\infty FF_N$).

We chose to state our main theorem for the case that $N \geq 2$ as the above theorem (and the rest of the paper) still applies in the case $N = 2$. However, as FF_2 is a Farey graph and $\partial_\infty FF_2$ is a Cantor set, we see that the result when $N = 2$ is not optimal.

1B. ***The intersection graph.*** A conjugacy class α of F_N is *geometric* if it is either part of a free basis of F_N, or else corresponds to the boundary curve of a once-holed surface with fundamental group identified with F_N. The *intersection graph* I_N (with Mann's definition [2014b], a variation on Kapovich and Lustig's [2009]) is the bipartite graph whose vertices are the simplicial F_N-trees in ∂cv_N together with the set of geometric conjugacy classes of F_N. A tree T is joined by an edge to a conjugacy class α whenever α is elliptic in T. Its hyperbolicity was proved in [Mann 2014b]. The intersection graph is also quasi-isometric to the Dowdall–Taylor cosurface graph [2017, Section 4]. We denote by $\mathcal{FAT}_N \subseteq \mathcal{AT}_N$ the space of free arational trees in $\overline{cv}_N$. Again, there is a coarsely Out(F_N)-equivariant map $\pi : cv_N \to I_N$, which sends a tree $T \in cv_N$ to a tree $\overline{T} \in \partial cv_N$ obtained by collapsing some of the edges of T to points.

Theorem 1.2 [Dowdall and Taylor 2017]. *There is a homeomorphism*

$$\partial\pi : \mathcal{FAT}_N/\!\sim\; \to \partial_\infty I_N,$$

which extends the map π continuously to the boundary.

Continuity of the extension is understood in the same way as in the statement of Theorem 1.1. When $N = 2$, the intersection graph I_2 is bounded, and $\partial_\infty I_2$ is empty.

1C. ***The cyclic splitting graph.*** A *cyclic splitting* of F_N is a simplicial, minimal F_N-tree in which all edge stabilizers are cyclic (possibly trivial). The *$\mathcal{Z}$-splitting graph* FZ_N is the graph whose vertices are the F_N-equivariant homeomorphism classes of $\mathcal{Z}$-splittings of F_N. Two splittings are joined by an edge if they have a common refinement. Its hyperbolicity was proved by Mann [2014a].

We recall that two trees $T, T' \in \overline{cv}_N$ are *compatible* if there exists a tree $\widehat{T} \in \overline{cv}_N$ that admits alignment-preserving F_N-equivariant maps onto both T and T'. A tree $T \in \overline{cv}_N$ is *$\mathcal{Z}$-averse* if it is not compatible with any tree $T' \in \overline{cv}_N$ which is compatible with a $\mathcal{Z}$-splitting of F_N. We denote by $\mathcal{X}_N$ the subspace of $\overline{cv}_N$ made of $\mathcal{Z}$-averse trees. Two $\mathcal{Z}$-averse trees are *equivalent* if they are both compatible with a common tree; although not obvious, this was shown in [Horbez 2016a] to be an equivalence relation on $\mathcal{X}_N$, which we denote by $\approx$.

There is an $\mathrm{Out}(F_N)$-equivariant map $\pi : cv_N \to FZ_N$, given by forgetting the metric.

Theorem 1.3 [Hamenstädt 2012; Horbez 2016a]. *There is a homeomorphism*

$$\partial\pi : \mathcal{X}_N/{\approx} \to \partial_\infty FZ_N,$$

which extends the map π continuously to the boundary.

Each $\approx$-equivalence class in $\mathcal{X}_N$ has preferred representatives that are mixing. We recall that a tree $T \in \overline{cv}_N$ is *mixing* if for all segments $I, J \subseteq T$, there exists a finite set $\{g_1, \dots, g_k\} \subseteq F_N$ such that J is contained in the union of finitely many translates $g_i I$.

Theorem 1.4 [Horbez 2016a]. *Every $\approx$-class in $\mathcal{X}_N$ contains a mixing tree, and any two mixing trees in the same $\approx$-class admit F_N-equivariant alignment-preserving bijections between each other. Any $\mathcal{Z}$-averse tree admits an F_N-equivariant alignment-preserving map onto every mixing tree in its $\approx$-class.*

The space of arational trees $\mathcal{AT}_N$ is contained in the space of $\mathcal{Z}$-averse trees $\mathcal{X}_N$, and the equivalence relation $\sim$ we have defined on $\mathcal{AT}_N$ is the restriction to $\mathcal{AT}_N$ of the equivalence relation $\approx$ on $\mathcal{X}_N$ (all arational trees are mixing [Reynolds 2012]). The inclusion $\mathcal{AT}_N \subseteq \mathcal{X}_N$ then induces a subspace inclusion $\partial_\infty FF_N \subseteq \partial_\infty FZ_N$.

1D. ***More on mixing $\mathcal{Z}$-averse trees.*** In this section, we establish a few more facts concerning mixing $\mathcal{Z}$-averse trees and their possible point stabilizers, building on the work in [Horbez 2016a]. If we were only concerned with the free factor graph, we would not need these results. The reader may decide to skim through the section and avoid the technicalities in the proofs in a first reading.

Lemma 1.5. *Let $T \in \overline{cv}_N$ be mixing and $\mathcal{Z}$-averse, and let $A \subseteq F_N$ be a proper free factor. Then the A-action on its minimal subtree $T_A \subseteq T$ is discrete* (*possibly T_A is* *reduced to a point*). *In particular*, *if no proper free factor is elliptic in T*, *then T is arational.*

Proof. Assume towards a contradiction that T_A is not simplicial. Since T has trivial arc stabilizers, the Levitt decomposition of T_A [1994] has trivial arc stabilizers (it may be reduced to a point in the case where T_A has dense orbits). Let $B \subseteq A$ be a free factor of A (hence of F_N) which is a vertex group of this decomposition. The B-minimal subtree T_B of T (which is also the B-minimal subtree of T_A) has dense orbits, so by [Reynolds 2011, Lemma 3.10] the family $\{g\overline{T_B} \mid g \in F_N\}$ is a transverse family in T. As T is mixing, it is a transverse covering. In addition, by [Reynolds 2012, Corollary 6.4] the stabilizer of T_B, and therefore the stabilizer of $\overline{T_B}$, is equal to B. But by [Horbez 2016a, Proposition 4.23], the stabilizer of a subtree in a transverse covering of a mixing $\mathcal{Z}$-averse tree cannot be a free factor (in fact, it cannot be elliptic in a $\mathcal{Z}$-splitting of F_N), so we get a contradiction. □

A collection $\mathcal{A}$ of subgroups of F_N is a *free factor system* if it coincides with the set of nontrivial point stabilizers in some simplicial F_N-tree with trivial arc stabilizers. There is a natural order on the collection of free factor systems, by saying that a free factor system $\mathcal{A}$ is *contained* in a free factor system $\mathcal{A}'$ whenever every factor in $\mathcal{A}$ is contained in one of the factors in $\mathcal{A}'$.

Proposition 1.6. *Let $T \in \overline{cv}_N$ be mixing and $\mathcal{Z}$-averse. Then either T is dual to an arational measured lamination on a closed hyperbolic surface with finitely many points removed*, *or else the collection of point stabilizers in T is a free factor system.*

Proof. First assume that there does not exist any free splitting of F_N in which all point stabilizers of T are elliptic. Then, with the terminology of [Horbez 2017, Section 5], the tree T is of surface type (by the same argument as in [Horbez 2017, Lemma 5.8]). Since T is $\mathcal{Z}$-averse, the skeleton of the dynamical decomposition of T is reduced to a point, in other words T is dual to an arational measured lamination on a surface.

Assume now that there exists a free splitting S of F_N in which all point stabilizers of T are elliptic. Let $\mathcal{A}$ be the smallest free factor system such that every point stabilizer of T is contained within some factor in $\mathcal{A}$. By Lemma 1.5, each factor A in $\mathcal{A}$ acts discretely on its minimal subtree T_A. In addition T_A has trivial arc stabilizers because T has trivial arc stabilizers. So either A is elliptic in T, or else T_A is a free splitting of A. The second situation cannot occur, as otherwise the point stabilizers in T_A would form a free factor system of A, contradicting the minimality of $\mathcal{A}$. Hence $\mathcal{A}$ coincides with the collection of point stabilizers of T. □

We will also establish a characterization of the mixing representatives in a given equivalence class of $\mathcal{Z}$-averse trees. Before that, we start with the following lemma.

Lemma 1.7. *Let $T, T' \in \overline{cv}_N$. Assume that T has dense orbits*, *and that there exists an F_N-equivariant alignment-preserving map $p : T \to T'$. Then either p is a* *bijection, or else there exists $g \in F_N$ that is elliptic in T' but not in T. In the latter case, some point stabilizer in T' is noncyclic.*

Proof. Notice that surjectivity of p follows from the minimality of T'. The collection of all subtrees of the form $p^{-1}(\{x\})$ with $x \in T'$ is a transverse family in T. If p is not a bijection, then one of the subtrees Y in this family is nondegenerate, and [Reynolds 2012, Proposition 7.6] implies that its stabilizer $A = \mathrm{Stab}(p(Y))$ is nontrivial, and that the A-action on Y is not discrete. This implies that the minimal A-invariant subtree of Y is not reduced to a point, and A contains an element that acts hyperbolically on T, while it fixes a point in T'. In addition A is not cyclic, so the last assertion of the lemma holds. □

Proposition 1.8. *A $\mathcal{Z}$-averse tree T is mixing if and only if for all $T' \approx T$, every element of F_N that is elliptic in T' is also elliptic in T. If T is dual to an arational lamination on a surface, then all trees in the $\approx$-class of T are mixing.*

Proof. If T is mixing, then all trees T' in its $\approx$-class admit an F_N-equivariant alignment-preserving map onto T, so every element elliptic in T' is also elliptic in T. If T is not mixing, then it admits an F_N-equivariant alignment-preserving map p onto a mixing tree $\overline{T}$ in the same $\approx$-class, and p is not a bijection. By Lemma 1.7, there exists an element of F_N that is elliptic in $\overline{T}$ but not in T. The second assertion in Proposition 1.8 follows from the last assertion of Lemma 1.7 because if T is dual to an arational lamination on a surface, then T is mixing and all point stabilizers in T are cyclic. □

Corollary 1.9. *Let T be a $\mathcal{Z}$-averse tree, and let $\mathcal{A}$ be a maximal free factor system elliptic in T. Let $\overline{T}$ be a mixing representative of the $\approx$-class of T. Then $\mathcal{A}$ is a maximal elliptic free factor system in $\overline{T}$ if and only if T is mixing.*

Proof. If T is mixing, then T and $\overline{T}$ have the same point stabilizers, so the conclusion is obvious. If T is not mixing, then Proposition 1.8 implies that $\overline{T}$ is not dual to a lamination on a surface, and that there exists an element $g \in F_N$ not contained in $\mathcal{A}$ that is elliptic in T. Proposition 1.6 shows that the collection of elliptic subgroups in $\overline{T}$ is a free factor system of F_N, and this free factor system strictly contains $\mathcal{A}$. □

2. Train-tracks, indices and stratifications

2A. *Train-tracks and carried trees.*

Definition 2.1 (train-tracks). A *train-track* τ is the data of

- a minimal, simplicial F_N-tree S^τ with trivial edge stabilizers,
- an equivalence relation $\sim_V^\tau$ on the set $V(S^\tau)$ of vertices of S^τ, such that if $v \sim_V^\tau v'$, then $gv \sim_V^\tau gv'$ for all $g \in F_N$, and such that no two adjacent vertices are equivalent,

- for each $\sim_V^\tau$-class X, an equivalence relation $\sim_{D,X}^\tau$ on the set $\mathcal{D}(X)$ of directions at the vertices in X, such that if two directions $d, d' \in \mathcal{D}(X)$ are equivalent, then for all $g \in F_N$, the directions $gd, gd' \in \mathcal{D}(gX)$ are also equivalent.

Equivalence classes of directions at X are called *gates* at X.

We denote by $\mathcal{A}(\tau)$ the free factor system made of all point stabilizers in S^τ.

Remark 2.2. Including an equivalence relation on the vertex set of S^τ in the definition may look surprising to the reader, as this is not standard in train-track theory for free groups. Roughly speaking, the equivalence classes of vertices correspond to branch points in trees carried by a track; this is explained in more detail in the Appendix.

We will usually also impose that the train-tracks we work with satisfy some additional assumptions. Let τ be a train-track. A pair (d, d') of directions based at a common vertex of S^τ is called a *turn*, and is said to be *legal* if d and d' are inequivalent. A subtree $A \subset S^\tau$ *crosses the turn* (d, d') if the intersection of A with both directions d and d' is nonempty. We say that A is *legal* in τ if each turn crossed by A is legal in τ.

Definition 2.3 (admissible train-tracks). A train-track τ is *admissible* if for every vertex $v \in S^\tau$, there exist three pairwise inequivalent directions d_1, d_2, d_3 at v such that for all $i \in \{1, 2, 3\}$, there exists an element $g_i \in F_N$ acting hyperbolically on S^τ whose axis in S^τ is legal and crosses the turn (d_i, d_{i+1}) (mod 3).

In particular, there are at least three gates at every equivalence class of vertices in admissible train-tracks. We will call a triple (g_1, g_2, g_3) as in Definition 2.3 a *tripod of legal elements* at v, and we say that their axes form a *tripod of legal axes* at v.

Edges in simplicial trees are given an affine structure, which enables us to consider maps from simplicial trees to $\mathbb{R}$-trees that are linear on edges. If $f : S \to T$ is linear then there is a unique metric on S for which f is isometric when restricted to every edge. We say that this is the *metric on S determined by the linear map f*.

Train-tracks naturally arise from morphisms between trees. Suppose that $f : S \to T$ is an F_N-equivariant map from a simplicial F_N-tree S to an F_N-tree T which is linear on edges and does not collapse any edge of S to a point. There is an induced equivalence relation $\sim_V$ on the set of vertices of S given by saying that two vertices are equivalent if they have the same f-image in T. Let w be a point in T and let $X := f^{-1}(w) \cap V(S)$ be its associated equivalence class in $V(S)$. As soon as $X \neq \emptyset$, equivariance of f implies that its (setwise) stabilizer Stab(X) in S is equal to the stabilizer of the point w in T. Let $\mathcal{D}(X)$ be the set of directions in S based at points in X. Since f does not collapse any edge, it induces an equivalence relation $\sim_{D,X}$ on $\mathcal{D}(X)$, where two directions in $\mathcal{D}(X)$ are equivalent if they have germs that map into the same direction at w in T. The set of Stab(X)-orbits of equivalence classes in $\mathcal{D}(X)$ then maps injectively into the set of $\mathrm{Stab}(w)$-orbits of directions at w in T. We call this collection of equivalence classes τ_f the *train-track structure on S induced by f*. This will be the typical example where a tree T is carried by the train-track τ_f. Our general definition of carrying is slightly more technical, as it involves a bit more flexibility with respect to the definition of f at exceptional classes of vertices, defined as follows.

Definition 2.4 (exceptional classes of vertices). Let τ be an admissible train-track. An equivalence class X of vertices of S^τ is called *exceptional* if there are exactly 3 gates at X.

Definition 2.5 (specialization). Let τ and τ' be two train-tracks, and let $v_0 \in S^\tau$ be such that $[v_0]$ is an exceptional equivalence class. We say that τ' is a *specialization* of τ at $[v_0]$ if

- $S^{\tau'} = S^\tau$,
- there exists a vertex $v_1 \in S^\tau$, not in the same orbit as v_0, such that $\sim_V^{\tau'}$ is the coarsest F_N-invariant equivalence relation finer than $\sim_V^\tau$ and such that $v_1 \sim_V^{\tau'} v_0$,
- if d, d' are two directions not based at any vertex in the orbit of $[v_0]$, then $d \sim^{\tau'} d'$ if and only if $d \sim^\tau d'$,
- every direction at a vertex in $[v_0]$ is equivalent to some direction at a vertex which is τ-equivalent to v_1.

Definition 2.6 (carrying). We say that an F_N-tree T is *carried* by a train-track τ if there is an F_N-equivariant map $f : S^\tau \to T$, which is linear on edges and does not collapse any edge to a point, such that the train-track structure τ_f induced by f is obtained from τ by a finite (possibly trivial) sequence of specializations.

In this situation, we call the map f a *carrying map* (with respect to τ).

Remark 2.7. The motivation for introducing specializations and allowing them in the definition of carrying is again explained in the Appendix; specializations will appear naturally later in the paper when we start performing folds on tracks.

Lemma 2.8. *The collection of all train-tracks τ for which there exists a tree $T \in \overline{cv}_N$ carried by τ is countable.*

Proof. There are countably many minimal, simplicial F_N-trees with trivial edge stabilizers. In addition, if τ carries a tree $T \in \overline{cv}_N$, then the stabilizer of every equivalence class X of vertices in τ is a finitely generated subgroup of F_N (because every point stabilizer in a very small F_N-tree is finitely generated by [Gaboriau and Levitt 1995]). The set $X/\mathrm{Stab}(X)$ is finite because there are finitely many F_N-orbits of vertices in S^τ, and if v, gv are two vertices in X, then $g \in \mathrm{Stab}(X)$. The stabilizer of every gate $[d]$ is at most cyclic, and again $[d]/\mathrm{Stab}([d])$ is finite. The equivalence relation on vertices is recovered by taking a finite set $A_1, \dots, A_k$ of representatives of F_N-orbits of equivalence classes, and in each A_i taking a finite set V_i of vertices such that $\mathrm{Stab}(A_i).V_i = A_i$. Each equivalence class is of the form $g\,\mathrm{Stab}(A_i).V_i$, for some $g \in F_N$. The equivalence relation on edges is recovered by taking a finite set $B_1, \dots, B_l$ of representatives of F_N-orbits of gates in τ, and for each B_i taking a finite set E_i of representatives of F_N-orbits of oriented edges determining the directions in B_i. Then each gate in τ is of the form $g\,\mathrm{Stab}(B_i).E_i$ for some $g \in F_N$. Hence τ is determined by the simplicial tree S^τ and the finite family $(\{\mathrm{Stab}\,(A_i)\}, \{V_i\}, \{\mathrm{Stab}(B_i)\}, \{E_i\})$, which gives a countable number of possible train-tracks. □

We will also need the following observation.

Lemma 2.9. *Let τ be an admissible train-track, and let $[a, b] \subseteq S^\tau$ be a legal segment. Then there exists an element $g \in F_N$ acting hyperbolically on S^τ, whose axis in S^τ is legal and contains $[a, b]$.*

Proof. Without loss of generality, we can assume that a and b are vertices of S^τ. Since τ is admissible, there exist elements $g, h \in F_N$ acting hyperbolically in S^τ, whose axes A_g, A_h are legal and pass through a, b respectively, only meet $[a, b]$ at one extremity, and such that the subtree $Y := A_g \cup [a, b] \cup A_h$ is legal. Standard properties of group actions on trees [Culler and Morgan 1987] imply that all turns in the axis of gh are contained in translates of Y. So the axis of gh is legal, and it contains $[a, b]$. □

The following two lemmas state that every tree $T \in \overline{cv}_N$ with trivial arc stabilizers is carried by an admissible train-track, and in addition the carrying map $f : S^\tau \to T$ is completely determined by the train-track structure. Given a free factor system $\mathcal{A}$, we say that a tree $T \in \overline{cv}_N$ is an *$\mathcal{A}$-tree* if all subgroups in $\mathcal{A}$ are elliptic in T.

Lemma 2.10. *Let $\mathcal{A}$ be a free factor system, and let $T \in \overline{cv}_N$ be an $\mathcal{A}$-tree with trivial arc stabilizers. Assume that $\mathcal{A}$ is a maximal free factor system such that T is an $\mathcal{A}$-tree. Then there exists an admissible train-track τ such that T is carried by τ and $\mathcal{A}(\tau) = \mathcal{A}$.*

Proof. We let $\mathcal{A}/F_N := \{[A_1], \dots, [A_k]\}$. If $\mathcal{A} \neq \varnothing$, then (up to replacing the subgroups A_i by appropriate conjugates) we can choose for S the universal cover of the graph of groups depicted in Figure 1. There is a unique F_N-equivariant map $f : S \to T$ that is linear on edges: indeed, every vertex v of S has nontrivial stabilizer G_v, and must be sent by f to the unique point in T fixed by G_v. The elements a_i and the subgroups $A_2, \dots, A_k$ cannot fix the same point as A_1 in T by maximality of $\mathcal{A}$, so f does not collapse any edge, and therefore τ_f is well-defined. Admissibility of τ_f can be checked by taking advantage of the fact that all vertex groups are infinite: one can construct the required legal elements by taking appropriate products of elliptic elements in T.

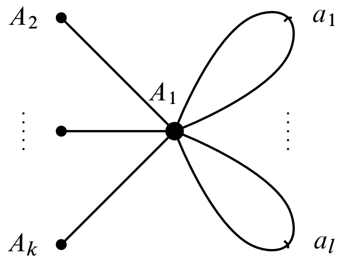


Figure 1. The tree S in the proof of Lemma 2.10, in the case where $\mathcal{A} \neq \varnothing$.

We now assume that $\mathcal{A} = \varnothing$. Let $\{a_1, \dots, a_N\}$ be a free basis of F_N such that $A = \langle a_1, a_2 \rangle$ acts freely with discrete orbits on T, and no a_i is elliptic in T. Such a basis exists as the reducing factors of T (if there are any) form a bounded subset of FF_N; see [Bestvina and Reynolds 2015, Corollary 5.3]. Let T_A be the minimal A-invariant subtree for the A-action on T. Let $\Gamma_A = T_A/A$. Then Γ_A is a finite graph in cv_2 with one or two vertices. We extend Γ_A to a marked F_N-graph Γ with the same number of vertices, by attaching loops labeled by the elements $a_3, \dots, a_N$ at a fixed vertex of Γ_A. Let S be the universal cover of Γ. The isometric embedding from T_A into T extends to a map $f : S \to T$, and f does not collapse any edges because no element a_i is elliptic in T. The induced train-track τ_f is admissible as each vertex of T_A has valence at least 3 and f is an F_N-equivariant isometric embedding when restricted to T_A. □

Lemma 2.11. *Let $T \in \overline{cv}_N$, and let τ be an admissible train-track such that T is carried by τ. Then there exists a unique carrying map $f : S^\tau \to T$ for the train-track structure. Furthermore, the carrying map f varies continuously on the set of trees that are carried by τ* (*in the equivariant Gromov–Hausdorff topology*).

Proof. To prove uniqueness, it is enough to show that the f-image of any vertex $v \in S^\tau$ is completely determined by the train-track structure. Let (g_1, g_2, g_3) be a tripod of legal elements at v, and let e_1, e_2, e_3 be the three edges at v taken by their axes. Then f is isometric when restricted to $e_1 \cup e_2 \cup e_3$, and also when restricted to the axes of g_1, g_2, g_3. This implies that the intersection of the axes of g_1, g_2, g_3 in T is reduced to a point, and f must send v to that point. Continuity also follows from the above argument. □

In order to define what it means for an equivalence class of $\mathcal{Z}$-averse trees to be carried by a train-track, we will need the following lemma.

Lemma 2.12. *Let $T, T' \in \overline{cv}_N$, and let τ be a train-track. If there is an F_N-equivariant alignment-preserving bijection from T to T', then T is carried by τ if and only if T' is carried by τ.*

Proof. Let $f : S^\tau \to T$ be a carrying map, and let $\theta : T \to T'$ be an F_N-equivariant alignment-preserving bijection. Let $f' : S^\tau \to T'$ be the unique linear map that coincides with $\theta \circ f$ on the vertices of S^τ. We claim that f' is a carrying map. Indeed, since θ is a bijection, two vertices in S^τ have the same f'-image if and only if they have the same f-image. In addition, since θ preserves alignment, the germs of two directions in S^τ are identified by f' if and only if they are identified by f. □

We recall that $\mathcal{X}_N \subseteq \overline{cv}_N$ denotes the subspace made of $\mathcal{Z}$-averse trees.

Definition 2.13 (carrying equivalence classes of $\mathcal{Z}$-averse trees). An equivalence class $\xi \in \mathcal{X}_N/{\approx}$ is *carried* by a train-track τ if some (equivalently, any) mixing representative of ξ is carried by τ.

2B. *Indices and stratifications.* We now define stratifications of the Gromov boundaries of I_N, FF_N and FZ_N by means of an index function taking finitely many values. We will define both the index of a train-track and the index of a tree $T \in \overline{cv}_N$, and explain how the two are related, before defining the index of a boundary point (which is an equivalence class of trees).

2B1. *Geometric index of a tree in $\overline{cv}_N$.* The following definition is reminiscent of the Gaboriau–Levitt index for trees in $\overline{cv}_N$ [1995], with a slight difference in the constants we use.

Definition 2.14 (geometric index of a tree $T \in \overline{cv}_N$). Let $T \in \overline{cv}_N$ be a tree with trivial arc stabilizers. The *geometric index* of T is defined as

$$i_{\mathrm{geom}}(T) := \sum_{v \in V(T)/F_N} (\alpha_v + 3\,\mathrm{rank}(\mathrm{Stab}(v)) - 3),$$

where $V(T)$ denotes the set of branch points in T, and α_v denotes the number of F_N-orbits of directions at v.

We warn the reader that the Gaboriau–Levitt index (denoted i_{GL}, below) is often also called the geometric index of a tree, and the definition used in this paper is somewhat nonstandard. The two are related in (1), below (see [Coulbois and Hilion 2012] for a discussion of various notions of index for trees and automorphisms).

Lemma 2.15. *For all $T \in \overline{cv}_N$ with trivial arc stabilizers, we have $i_{\mathrm{geom}}(T) \le 3N-3$. If T is arational, then $i_{\mathrm{geom}}(T) \le 2N-2$. If T is free and arational, then $i_{\mathrm{geom}}(T) \le 2N-3$.*

Proof. Let

$$i_{GL}(T) := \sum_{v \in V(T)/F_N} (\alpha_v + 2\,\mathrm{rank}(\mathrm{Stab}(v)) - 2).$$

We have $i_{\text{geom}}(T) \leq \frac{3}{2} i_{GL}(T)$ for all $T \in \overline{cv}_N$, and Gaboriau and Levitt [1995, Theorem III.2] proved that $i_{GL}(T) \leq 2N-2$, so $i_{\text{geom}}(T) \leq 3N-3$. In addition,

$$i_{\text{geom}}(T) = i_{GL}(T) + \sum_{v \in V(T)/F_N} (\text{rank}(\text{Stab}(v)) - 1). \tag{1}$$

Assume now that T is arational. If the F_N-action on T is free, since there is at least one orbit of branch points in T, we get from (1) that $i_{\text{geom}}(T) < i_{GL}(T) \leq 2N-2$. Otherwise, by [Reynolds 2012], the F_N-action on T is dual to an arational measured lamination on a once-holed surface, so all point stabilizers in T are either trivial or cyclic, and we get $i_{\text{geom}}(T) \leq i_{GL}(T) \leq 2N-2$. □

2B2. *Index of a train-track and carrying index of a tree.* The *height* $h(\mathcal{A})$ of a free factor system $\mathcal{A}$ is defined as the maximal length k of a proper chain of free factor systems $\varnothing = \mathcal{A}_0 \subsetneq \mathcal{A}_1 \subsetneq \cdots \subsetneq \mathcal{A}_k = \mathcal{A}$. We recall that the free factor system $\mathcal{A}(\tau)$ associated to a train-track τ is the free factor system consisting of the vertex stabilizers in the associated tree S^τ.

Definition 2.16 (index of a train-track). The *geometric index* of a train-track τ is defined as

$$i_{\text{geom}}(\tau) := \sum_{i=1}^{k} (\alpha_i + 3r_i - 3),$$

where the sum is taken over a finite set $\{X_1, \ldots, X_k\}$ of representatives of the F_N-orbits of equivalence classes in $V(S^\tau)$, α_i denotes the number of $\text{Stab}(X_i)$-orbits of gates at the vertices in X_i, and r_i denotes the rank of $\text{Stab}(X_i)$.

The *index* of τ is defined to be the pair $i(\tau) := (h(\mathcal{A}(\tau)), i_{\text{geom}}(\tau))$.

Indices of train-tracks will be ordered lexicographically.

Remark 2.17. If τ is admissible, then every equivalence class X of vertices in τ has a nonnegative contribution to the geometric index of τ. The contribution of a class X to the geometric index of τ is zero if and only if $\text{Stab}(X)$ is trivial, and there are exactly three gates at X; if τ carries a tree with trivial arc stabilizers then this is precisely when X is exceptional.

Lemma 2.18. *Let $T \in \overline{cv}_N$ be a tree with trivial arc stabilizers, and let τ be an admissible train-track that carries T. Then $i_{\text{geom}}(\tau) \leq i_{\text{geom}}(T)$.*

Proof. Let $f : S^\tau \to T$ be the unique carrying map. Since τ is admissible, vertices in S^τ are mapped to branch points in T. As f is a carrying map, distinct equivalence classes of nonexceptional vertices in S^τ are mapped to distinct branch points in T. If X is an equivalence class in S^τ mapping to a branch point $v \in T$ then $\text{Stab}(X) = \text{Stab}(v)$. Furthermore, one checks that two gates based at X in distinct $\text{Stab}(X)$-orbits are mapped under f to directions at v in distinct $\text{Stab}(v)$-orbits.

Since exceptional vertices do not contribute to the geometric index, it follows that $i_{\text{geom}}(\tau) \leq i_{\text{geom}}(T)$. □

Notice that the inequality from Lemma 2.18 might be strict if some branch direction in the tree T is not "visible" in the track. In the following definition, instead of directly counting branch points and branch directions in T, we will count the maximal number of such directions that are visible from a train-track.

Definition 2.19 (carrying index of a tree $T \in \overline{cv}_N$). Let $T \in \overline{cv}_N$ be a tree with trivial arc stabilizers. We define the *carrying index* of T, denoted by $i(T)$, as the maximal index of an admissible train-track τ that carries T. An *ideal carrier* for T is an admissible train-track τ that carries T such that $i(T) = i(\tau)$.

Remark 2.20. In view of Lemma 2.10, if τ is an ideal carrier of a tree T, then $\mathcal{A}(\tau)$ is a maximal free factor system that is elliptic in T.

Lemma 2.21. *The carrying index of a tree $T \in \overline{cv}_N$ with trivial arc stabilizers can only take boundedly many values* (*with a bound only depending on N*). *The carrying index of an arational tree is comprised between* $(0, 0)$ *and* $(0, 2N-2)$. *The carrying index of a free arational tree is comprised between* $(0, 0)$ *and* $(0, 2N-3)$.

Proof. The first assertion is a consequence of Lemma 2.15, together with the fact that there is a bound (only depending on N) on the height of a free factor system of F_N. The other assertions follow from Lemma 2.15 because no nontrivial free factor is elliptic in an arational tree, so the height $h(\mathcal{A}(\tau))$ for a train-track τ that carries an arational tree is 0. □

2B3. *Index of a boundary point and stratifications.* We now define the index of an equivalence class ξ of $\mathcal{Z}$-averse trees. We recall that by definition a train-track τ carries ξ if and only if τ carries the mixing representatives of ξ.

Definition 2.22 (index of a boundary point). The *index* $i(\xi)$ of an equivalence class $\xi \in \mathcal{X}_N/{\approx}$ is defined as the maximal index of an admissible train-track that carries ξ (equivalently $i(\xi)$ is the carrying index of the mixing representatives of ξ).

Given an admissible train-track τ, we define the *cell* $P(\tau)$ as the subspace of $\mathcal{X}_N/{\approx}$ made of all classes ξ that are carried by τ and such that $i(\xi) = i(\tau)$.

For all i, we define the *stratum* $X_i \subseteq \mathcal{X}_N/{\approx}$ as the set of all points $\xi \in \mathcal{X}_N/{\approx}$ such that $i(\xi) = i$.

The Gromov boundary $\partial_\infty FZ_N$ can be written as the union of all strata X_i, where i varies over the finite set of all possible indices for mixing $\mathcal{Z}$-averse trees.

If a tree $T \in \overline{cv}_N$ is mixing and $\mathcal{Z}$-averse, and if no proper free factor is elliptic in T, then Lemma 1.5 says that T is arational. Therefore, the boundary $\partial_\infty FF_N$ coincides with the subspace of $\mathcal{X}_N/{\approx}$ which is the union of all strata X_i with i comprised between $(0, 0)$ and $(0, 2N-2)$.

Finally, the boundary $\partial_\infty I_N$ can be written as the union $X'_0 \cup \cdots \cup X'_{2N-3}$, where X'_i is the subspace of $X_{(0,i)}$ made of equivalence classes of free actions.

In view of the topological facts recalled in the introduction, we are left showing that each cell $P(\tau)$ is closed in its stratum $X_{i(\tau)}$, and that $\dim(P(\tau)) \leq 0$. This will be the contents of Sections 3 and 4, respectively. We give a complete overview of the proof of our main theorem in Section 5.

3. Closedness of $P(\tau)$ in its stratum

In general, the property of a tree being carried by a train-track is not a closed condition. However, cells determined by train-tracks are closed in their own strata. This is the goal of the present section. As each boundary we study is metrizable, throughout the paper we will use sequential arguments to work with the topology where this is appropriate.

Proposition 3.1. *Let τ be an admissible train-track. Then all points in $\partial P(\tau) = \overline{P(\tau)} \setminus P(\tau) \subseteq \mathcal{X}_N/{\approx}$ have index strictly greater than $i(\tau)$. In particular $P(\tau)$ is closed in $X_{i(\tau)}$.*

Our proof of Proposition 3.1 is based on Lemma 3.2 and Proposition 3.4 below.

Lemma 3.2. *Let τ be an admissible train-track, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of mixing $\mathcal{Z}$-averse trees carried by τ. Assume that the trees T_n converge to a $\mathcal{Z}$-averse tree T, and denote by ξ the $\approx$-class of T. Then either T is mixing and $\mathcal{A}(\tau)$ is a maximal free factor system that is elliptic in T, or else $i(\xi) > i(\tau)$.*

Proof. The free factor system $\mathcal{A}(\tau)$ is elliptic in all trees T_n, and therefore it is also elliptic in T. If $\mathcal{A}(\tau)$ is not a maximal free factor system elliptic in T, then as T collapses to any mixing representative $\overline{T} \in \xi$ the free factor system $\mathcal{A}(\tau)$ is not maximally elliptic in $\overline{T}$. Remark 2.20 then implies that $i(\xi) > i(\tau)$. If T is not mixing (but possibly $\mathcal{A}(\tau)$ is maximally elliptic in T), then Corollary 1.9 implies that again $\mathcal{A}(\tau)$ is not maximally elliptic in any mixing representative of ξ. Hence $i(\xi) > i(\tau)$ in this case, also. □

To complete the proof of Proposition 3.1, we are thus left understanding the case where the limiting tree T is mixing, and $\mathcal{A}(\tau)$ is a maximal elliptic free factor system in T: this will be done in Proposition 3.4 below. The idea is that the carrying maps $f_n : S^\tau \to T_n$ will always converge to an F_N-equivariant map $f : S^\tau \to T$; in general the limiting map f can fail to be a carrying map (it may even collapse some edges to points), and in this case we will prove in Proposition 3.4 that there is a jump in index. We start by proving the existence of the limiting map f.

Lemma 3.3. *Let τ be an admissible train-track, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of trees in $\overline{cv}_N$ with trivial arc stabilizers, converging to a tree $T \in \overline{cv}_N$. Assume that all trees T_n are carried by τ, and for all $n \in \mathbb{N}$, let $f_n : S^\tau \to T_n$ be the carrying* *map. Then the maps f_n converge* (*in the equivariant Gromov–Hausdorff topology*) *to a map $f : S^\tau \to T$.*

Proof. Let v be a vertex in S^τ. Since τ is admissible, there exist three inequivalent edges e_1, e_2, e_3 that form a legal tripod at v, and two elements $g, h \in F_N$ that both act hyperbolically in S^τ, whose axes are legal, and such that the axis of g (resp. h) crosses the turn (e_1, e_3) (resp. (e_2, e_3)). Then the axes of g and h in S^τ intersect in a compact nondegenerate segment with initial point v, and up to replacing g and h by their inverses, we can assume that g and h both translate along this segment in the direction going out of v. Since all maps f_n are carrying maps, the elements g and h are hyperbolic in all trees T_n, and their axes in T_n intersect in a compact nondegenerate segment, on which they translate in the same direction. In the limiting tree T the intersection $A_g^T \cap A_h^T$ of the axes (or fixed sets) of g and h is still a compact segment, and if it is nondegenerate the elements g and h are hyperbolic in T and still translate in the same direction along the intersection. We then define $f(v)$ to be the initial point of $A_g^T \cap A_h^T$. We repeat this process over each orbit of vertices to obtain an F_N-equivariant map from the vertices of S^τ to T, and extend the map linearly over edges. Distances between intersections of axes (as well as their initial points) are determined by the Gromov–Hausdorff topology [Paulin 1989], so it follows that for any two vertices v and v', the distance $d_{T_n}(f_n(v), f_n(v'))$ converges to $d_T(f(v), f(v'))$ as n goes to infinity. This implies that the sequence of maps (f_n) converges to f. □

Proposition 3.4. *Let τ be an admissible train-track, and let $(T_n)_{n\in\mathbb{N}}$ be a sequence of mixing $\mathcal{Z}$-averse trees carried by τ. Assume that (T_n) converges to a mixing $\mathcal{Z}$-averse tree T, and that $\mathcal{A}(\tau)$ is a maximal free factor system elliptic in T. For all $n \in \mathbb{N}$, let $f_n : S^\tau \to T_n$ be the carrying map, and let $f : S^\tau \to T$ be the limit of the maps f_n. Let S' be the tree obtained from S^τ by collapsing all edges whose f-image is reduced to a point, let $f' : S' \to T$ be the induced map, and let $\tau' := \tau_{f'}$ be the train-track on S' induced by f'. Then τ' is admissible. Either $i_{\mathrm{geom}}(\tau') > i_{\mathrm{geom}}(\tau)$, or else $S' = S^\tau$ and τ' is obtained from τ by a finite* (*possibly trivial*) *sequence of specializations* (*hence τ carries T*).

Before proving Proposition 3.4, we first complete the proof of Proposition 3.1 from the above facts.

Proof of Proposition 3.1. Let $\xi \in \partial P(\tau)$. If ξ is carried by τ, then $i(\xi) \geq i(\tau)$, and this inequality is strict because otherwise τ would be an ideal carrier of ξ, contradicting $\xi \in \partial P(\tau)$.

We now assume that ξ is not carried by τ. Let $(\xi_n)_{n\in\mathbb{N}} \in P(\tau)^{\mathbb{N}}$ be a sequence converging to ξ, and for all $n \in \mathbb{N}$, let T_n be a mixing representative of the equivalence class ξ_n. Since $\overline{cv}_N$ is projectively compact, up to a subsequence we can assume that $(T_n)_{n\in\mathbb{N}}$ converges to a tree T. Since the boundary map $\partial\pi : \mathcal{X}_N \to \partial_\infty FZ_N$ is closed, the tree T is $\mathcal{Z}$-averse, and in the $\approx$-class ξ. Using Lemma 3.2, the proof reduces to the case where T is mixing, and $\mathcal{A}(\tau)$ is a maximal free factor system elliptic in T. Since ξ is not carried by τ, Proposition 3.4 then shows that T is carried by a train-track τ' satisfying $i_{\text{geom}}(\tau') > i_{\text{geom}}(\tau)$. This implies in turn that $i(\xi) > i(\tau)$. $\square$

The rest of the section is devoted to the proof of Proposition 3.4.

Proof of Proposition 3.4.
1. *The track τ' is admissible.* Before proving that τ' is admissible, we first observe that all hyperbolic elements in S^τ are still hyperbolic in S'. Indeed, in view of Proposition 1.6 and of our assumption that $\mathcal{A}(\tau)$ is a maximal elliptic free factor system in T, if $g \in F_N \setminus \{e\}$ is elliptic in T but is not contained in a free factor from $\mathcal{A}(\tau)$, then the conjugacy class of g is given by (some power of) a boundary curve. Hence g is not contained in any proper free factor of F_N relative to $\mathcal{A}(\tau)$. Therefore the axis of g in S^τ crosses all orbits of edges, and so cannot be collapsed to a point by the collapse map $\pi : S^\tau \to S'$.

We now prove that τ' is admissible. Let $v \in V(S')$, and let $X \subseteq S^\tau$ be the π-preimage of v. We first observe that X is a bounded subtree of S^τ: indeed, otherwise, we would find two oriented edges e, e' in X in the same F_N-orbit (say $e' = ge$) and pointing in the same direction; this would imply that g is hyperbolic in S^τ but not in S', a contradiction. Let $Y \subseteq X$ be a maximal legal subtree of X. Using the fact that τ is admissible, we can find three pairwise inequivalent edges e_1, e_2, e_3 lying outside of X based at extremal vertices v_1, v_2, v_3 of Y (these vertices are not necessarily distinct), such that the subtree $Y \cup e_1 \cup e_2 \cup e_3$ is legal in τ. Furthermore, no edge e_i is collapsed to a point by π. Since τ is admissible, Lemma 2.9 gives three elements g_1, g_2, g_3, which act hyperbolically in S^τ, whose axes are legal, and such that the axis of g_i in S^τ crosses the segment $e_i \cup [v_i, v_{i+1}] \cup e_{i+1}$ (mod 3). For all $n \in \mathbb{N}$, the map f_n preserves alignment when restricted to each of these axes, so in the limit f also preserves alignment when restricted to these axes. This implies that g_1, g_2 and g_3 are legal in S'. In addition, their axes form a legal tripod at v, so τ' is admissible.

2. *Controlling the index of τ'.* We now prove that $i_{\text{geom}}(\tau') > i_{\text{geom}}(\tau)$ unless τ carries T. Given a vertex $v \in V(S')$, we will first establish that the contribution of the equivalence class $[v]$ to the index of τ' is no less than the sum of the contributions of its π-preimages in $V(S^\tau)$ to the index of S^τ.

Let $x := f'(v)$. Let X be the set of vertices in S^τ that are mapped to x under f. As equivalent vertices in τ are mapped to the same point in T under f, the set X is a union of equivalence classes of vertices in τ. As $\mathrm{Stab}(x) = \mathrm{Stab}(X)$, if two vertices in X are in the same F_N-orbit, then they are in the same $\mathrm{Stab}(x)$-orbit. Hence we may pick a finite set $E_1, \dots, E_k$ of representatives of the $\mathrm{Stab}(x)$-orbits of equivalence classes of vertices in X. Suppose that the images of these equivalence classes correspond to l orbits of points in T_n (where l can be chosen to be independent of n by passing to an appropriate subsequence). After reordering (and possibly passing to a further subsequence) we may assume $E_1, \dots, E_l$ are mapped to distinct Stab(x)-orbits of points in T_n for all $n \in \mathbb{N}$, and $E_{l+1}, \dots, E_k$ are exceptional classes in τ (notice that some of the classes E_i with $i \in \{1, \dots, l\}$ might be exceptional as well).

Suppose that each equivalence class E_i has α_i gates in τ and has a stabilizer of rank r_i. As exceptional classes do not contribute to the geometric index, the amount $E_1, \dots, E_k$ contribute to the geometric index of τ is

$$\sum_{i=1}^{l} \alpha_i + 3 \sum_{i=1}^{l} r_i - 3l. \tag{2}$$

Let $E = [v]$ be the equivalence class corresponding to the image of X in τ'. Then E contributes $\alpha' + 3r' - 3$ to the index of τ', where α' is the number of orbits of gates at E and r' is the rank of Stab(x). If we define $\alpha = \sum_{i=1}^{l} \alpha_i$ and $r = \sum_{i=1}^{l} r_i$, then the difference between the index contribution of E in τ' and the index contribution of $E_1, \dots, E_k$ in τ is

$$\alpha' - \alpha + 3(r' - r) + 3(l - 1). \tag{3}$$

Our goal is to control the number of Stab(x)-orbits of gates lost when passing from τ to τ'. For all $n \in \mathbb{N}$, we let Y_n be the subtree of T_n spanned by the points in $f_n(X)$. Since there are finitely many Stab(x)-orbits of equivalence classes of vertices in X, the tree Y_n is obtained from the Stab(x)-minimal invariant subtree of T_n by attaching finitely many orbits of finite trees (if Stab(x) $= \{e\}$, then Y_n is a finite subtree of T_n). Recall that we assumed that $\mathcal{A}(\tau)$ is a maximal elliptic free factor system in T. Using Proposition 1.6, this implies that Stab(x) is either cyclic or contained in $\mathcal{A}(\tau)$ (it may be trivial). In the second case Stab(x) is elliptic in T_n. Hence the quotient $G_n = Y_n / \mathrm{Stab}(x)$ is a finite graph of rank 0 or 1. There are l marked points in G_n corresponding to the images of $E_1, \dots, E_l$.

Claim 1. *If e is an oriented edge in S^τ which is collapsed under π, and such that $f'(\pi(e)) = x$, then the gate corresponding to e in τ is mapped under f_n to a direction in Y_n at a point in $f_n(X)$.*

Claim 2. *If e and e' are oriented edges in S^τ that determine inequivalent directions d, d' in τ, based at vertices in* Stab(x).$\{E_1, \dots, E_l\}$, *and if $\pi(e)$ and $\pi(e')$ are nondegenerate and equivalent in τ', then for all sufficiently large $n \in \mathbb{N}$, one of the directions d, d' is mapped under f_n to a direction in Y_n at a point in $f_n(X)$.*

Proof of claims. For Claim 1, as e is collapsed its endpoints lie in X, so that $f_n(e)$ is an arc between two points of $f_n(X)$ in Y_n. For Claim 2, if $\pi(e)$ and $\pi(e')$ are equivalent, then $f(e) \cap f(e')$ is nondegenerate. Then for all sufficiently large $n \in \mathbb{N}$, the intersection of $f_n(e)$ with $f_n(e')$ contains a nondegenerate segment. As e and e' are not equivalent in τ, they are based at distinct equivalence classes of vertices in $\mathrm{Stab}(x).\{E_1, \dots, E_l\}$. Let us call these $[v_1]$ and $[v_2]$. We have $f_n(v_1) \neq f_n(v_2)$, and the arc connecting $f_n(v_1)$ with $f_n(v_2)$ is contained in the tree Y_n, and is covered by the union of the two arcs corresponding to $f_n(e)$ (starting at $f_n(v_1)$) and $f_n(e')$ (starting at $f_n(v_2)$). One can check that the initial direction of either $f_n(e)$ or $f_n(e')$ must then be contained in Y_n.

Since T has trivial arc stabilizers, no two oriented edges in the same orbit are equivalent in τ', and therefore there are only finitely many orbits of pairs of equivalent directions in τ'. Therefore, we can choose $n \in \mathbb{N}$ large enough so that the conclusion of Claim 2 holds for all pairs of inequivalent directions in S^τ based at vertices in $\mathrm{Stab}(x).\{E_1, \dots, E_l\}$, whose π-images are (nondegenerate and) equivalent in S'.

Denote by $\mathcal{G}_n^{\mathrm{out}}$ (resp. $\mathcal{G}_n^{\mathrm{in}}$) the set of gates based at vertices in $\mathrm{Stab}(x).\{E_1, \dots, E_l\}$ which are mapped outside of Y_n (resp. inside Y_n) by f_n. Claim 1 implies that there is a map Ψ from the set E_n^{out} of edges in $\mathcal{G}_n^{\mathrm{out}}$, to the set of gates at E in τ'. Claim 2 then implies that any two edges in E_n^{out} in distinct gates have distinct Ψ-images, so $|\mathcal{G}_n^{\mathrm{out}}/\,\mathrm{Stab}(x)| \leq \alpha'$. Furthermore, our definition of $E_1, \dots, E_l$ implies that the natural map (induced by f_n) from $\mathcal{G}_n^{\mathrm{in}}/\,\mathrm{Stab}(x)$ to the set $\mathrm{Dir}(G_n)$ of directions based at the marked points in the quotient graph G_n, is injective, so $|\mathcal{G}_n^{\mathrm{in}}/\,\mathrm{Stab}(x)| \leq |\mathrm{Dir}(G_n)|$. By summing the above two inequalities, it follows that $\alpha - \alpha'$ is bounded above by the number of directions based at the marked points in the quotient graph G_n.

We will now distinguish three cases. Notice that up to passing to a subsequence, we can assume that one of them occurs.

Case 1: The stabilizer $\mathrm{Stab}(x)$ is either trivial, or fixes a point x_n in all trees T_n, and for all $n \in \mathbb{N}$, there is a vertex $v_n \in V(S^\tau)$ such that $f_n(v_n) = x_n$ (this happens in particular if $\mathrm{Stab}(x)$ is contained in $\mathcal{A}(\tau)$).

In this case, we have $r = r'$, and the graph G_n has rank 0. The l marked points of G_n include the leaves of G_n: indeed, every leaf of G_n is either the projection to G_n of a point in $f_n(X)$, or else it is the projection to G_n of the unique point $y \in T_n$ with stabilizer equal to $\mathrm{Stab}(x)$. In the latter case, we again have $y \in f_n(X)$ by assumption. We will apply the following fact to the graph G_n.

Fact. *If G is a finite connected graph of rank r with l marked points containing all leaves of G, then there are at most $2(l + r - 1)$ directions at these marked points.*

Proof of the fact. An Euler characteristic argument shows that in any finite connected graph G with v vertices, there are exactly $2(v + r - 1)$ directions at the vertices. Viewing the vertex set as a set of marked points, removing any nonleaf vertex from this set loses at least two directions, from which the fact follows. □

This fact applied to the quotient graph G_n shows that

$$\alpha' - \alpha \geq -2(l-1). \tag{4}$$

From (3) we see that the new equivalence class contributes at least $l-1$ more to the geometric index than the sum of the contributions of the previous equivalence classes to the index of τ. This shows that the index increases as soon as $l \geq 2$, so we are left with the case where $l = 1$.

In this remaining situation, the graph G_n is a single point, so all but possibly one (say E_1) of the equivalence classes $E_1, \dots, E_k$ are exceptional. By Claim 1 no edges corresponding to directions at X are collapsed under π. Furthermore, by Claim 2 distinct equivalence classes of gates based at vertices in the orbit of E_1 are mapped to distinct gates under π. It follows that either there are more gates at the equivalence class X when passing from τ to τ' (so the index increases), or the gates corresponding to directions at E_1 are mapped bijectively and each exceptional class in $E_2, \dots, E_k$ corresponds to a specialization of τ (otherwise we would see extra gates). Repeating this argument across all equivalence classes in τ', we find that either the geometric index of τ' is greater than that of τ, or τ' is obtained from τ by applying a finite number of specializations. In this latter case, this implies that T is carried by τ.

Case 2: The group Stab(x) is not elliptic in T_n (in particular Stab(x) is cyclic).

In this case, we have $r' = 1$ and $r = 0$. The graph G_n is then a circle, with finitely many finite trees attached, whose leaves are the projections of points in $f_n(X)$. In particular G_n has rank $r' - r = 1$, and all its leaves are marked. By applying the above fact, we thus get that

$$\alpha' - \alpha \geq -2(l + r' - r - 1). \tag{5}$$

Since $r' - r = 1$, we get from (3) that the new equivalence class contributes at least l more to the geometric index than the sum of the contributions of the previous equivalence classes to the index of τ, so $i(\tau') > i(\tau)$.

Case 3: The stabilizer Stab(x) is cyclic, fixes a point x_n in all trees T_n, but no vertex in S^τ is mapped to x_n under f_n.

In this case, we have $r = 0$ and $r' = 1$. The graph G_n has rank 0. The image of x_n in the quotient graph G_n might not be marked, however the set of marked points in G_n includes all other leaves. We will use the following variation on the above fact (which is easily proved by first including the missing leaf to the set of marked points).

Fact. *Suppose G is a finite connected graph of rank r with a set of l marked points containing all leaves except possibly one. Then there are at most $2(l + r - 1) + 1$ directions at these marked points.*

This fact, applied to the graph G_n, shows that

$$\alpha' - \alpha \geq -2(l-1) - 1. \tag{6}$$

Since $r' - r = 1$, we get from (3) that the new equivalence class contributes at least $l+1$ more to the geometric index than the sum of the contributions of the previous equivalence classes to the index of τ, and again we are done. □

4. The cells $P(\tau)$ have dimension at most 0.

The goal of this section is to prove the following fact.

Proposition 4.1. *Let τ be an admissible train-track. Then $P(\tau)$ has dimension at most* 0.

The strategy of our proof of Proposition 4.1 is the following. Given any point $\xi \in P(\tau)$, our goal is to construct arbitrarily small open neighborhoods of ξ with empty boundary in $P(\tau)$. This will be done using folding sequences of train-tracks. In Section 2, we defined the notion of a specialization, which gives a new train-track τ' from a train-track τ. In Section 4A, we will introduce other operations called *folding moves*. These will enable us to define a new train-track τ' from a train-track τ by folding at an illegal turn. We will then make the following definition.

Definition 4.2 (folding sequence of train-tracks). A *folding sequence of train-tracks* is an infinite sequence $(\tau_i)_{i\in\mathbb{N}}$ of admissible train-tracks such that for all $i \in \mathbb{N}$, the train-track τ_{i+1} is obtained from τ_i by applying a folding move followed by a finite (possibly trivial) sequence of specializations.

Given $\xi \in \partial_\infty FZ_N$, we say that the folding sequence of train-tracks $(\tau_i)_{i\in\mathbb{N}}$ is *directed by* ξ if $\xi \in P(\tau_i)$ for all $i \in \mathbb{N}$.

We will first prove in Section 4A that folding sequences of train-tracks exist.

Lemma 4.3. *Let τ be an admissible train-track, and let $\xi \in P(\tau)$. Then there exists a folding sequence of train-tracks $(\tau_i)_{i\in\mathbb{N}}$ directed by ξ, such that τ_0 is obtained from τ by a finite* (*possibly trivial*) *sequence of specializations.*

We will then prove in Section 4B that all sets $P(\tau_i)$ in a folding sequence of train-tracks are open in $P(\tau_0)$, by showing the following two facts.

Lemma 4.4. *Let τ be an admissible train-track, and let τ' be a specialization of τ. Then τ' is admissible, and $P(\tau')$ is an open subset of $P(\tau)$.*

Lemma 4.5. *Let τ be an admissible train-track, and let τ' be a train-track obtained from τ by folding an illegal turn. Then τ' is admissible, and $P(\tau')$ is an open subset of $P(\tau)$.*

In other words, the sets $P(\tau_i)$ are open neighborhoods of ξ in $P(\tau)$, and they are closed in $P(\tau)$ (because their boundaries are made of trees with higher index in view of Proposition 3.1). So to complete the proof of Proposition 4.1, we are left showing that $P(\tau_i)$ can be made arbitrary small. This is proved in Section 4C in the form of the following proposition.

Lemma 4.6. *Let τ be an admissible train-track, let $\xi \in P(\tau)$, and let $(\tau_i)_{i\in\mathbb{N}}$ be a folding sequence of train-tracks directed by ξ. Then the diameter of $P(\tau_i)$ converges to* 0.

We now sum up the proof of Proposition 4.1 from the above four lemmas.

Proof of Proposition 4.1. Assume that $P(\tau) \neq \emptyset$, and let $\xi \in P(\tau)$. Let $\varepsilon > 0$. Let $(\tau_i)_{i\in\mathbb{N}}$ be a folding sequence of train-tracks directed by ξ provided by Lemma 4.3, where τ_0 is obtained from τ by a finite (possibly trivial) sequence of specializations. Lemma 4.6 shows that we can find $k \in \mathbb{N}$ such that $P(\tau_k)$ contains ξ and has diameter at most ε. An iterative application of Lemmas 4.4 and 4.5 then ensures that $P(\tau_k)$ is an open neighborhood of ξ in $P(\tau)$. Furthermore, the boundary of $P(\tau_k)$ in $P(\tau)$ is empty (it is made of trees of higher index by Proposition 3.1). As each point in $P(\tau)$ has arbitrarily small open neighborhoods with empty boundary, $\dim(P(\tau)) = 0$. □

4A. *Existence of folding sequences of train-tracks directed by a boundary point.*

4A1. *More on specializations.* The notion of specialization is given in Definition 2.5.

Lemma 4.7. *Let τ be an admissible train-track, and let τ' be a specialization of τ. Then τ' is admissible, and $i(\tau') = i(\tau)$.*

Proof. Admissibility is clear, as the underlying tree S^τ and the collection of legal turns at each vertex of S^τ are unchanged after a specialization. To see that $i(\tau') = i(\tau)$, notice first that the exceptional vertices in S^τ contribute 0 to the geometric index of τ. Stabilizers do not increase by definition of a specialization (otherwise two inequivalent vertices in τ that are not in the orbit of v_0 would become equivalent in τ'), and no new gate is created. So $i_{\text{geom}}(\tau') = i_{\text{geom}}(\tau)$, and since stabilizers of equivalence classes of vertices are the same in τ and in τ', we have $i(\tau') = i(\tau)$. □

4A2. *Folding moves.* We now introduce three types of folding moves and discuss some of their properties.

Definition 4.8 (singular fold, see Figure 2). Let τ be a train-track. Let $e_1 = [v, v_1]$ and $e_2 = [v, v_2]$ be two edges in S^τ that are based at a common vertex v, determine equivalent directions at v, and such that $v_1 \sim^\tau v_2$.

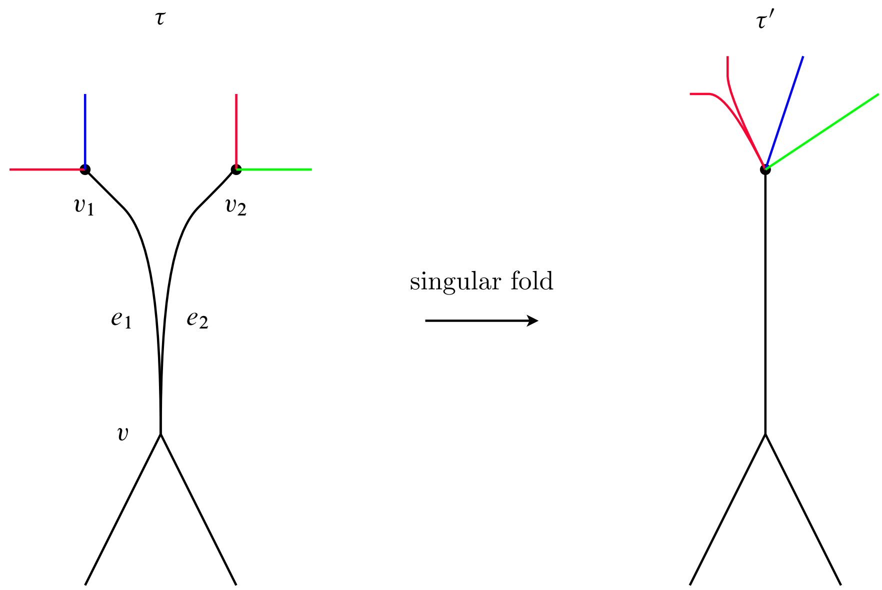


Figure 2. A singular fold. The two black vertices v_1 and v_2 are equivalent in τ, and the colors give the gates at the equivalence class of vertices they define.

A train-track τ' is obtained from τ by a *singular fold* at $\{e_1, e_2\}$ if

- there is an F_N-equivariant map $g : S^\tau \to S^{\tau'}$ that consists in equivariantly identifying e_1 with e_2,
- for all vertices $v, v' \in V(S^\tau)$, we have $g(v') \sim^{\tau'} g(v)$ if and only if $v' \sim^\tau v$,
- for all directions d, d' in S^τ based at equivalent vertices of S^τ, we have $g(d') \sim^{\tau'} g(d)$ if and only if $d' \sim^\tau d$.

Remark 4.9. If τ satisfies the hypothesis of Definition 4.8 with edges e_1 and e_2, then such a singular fold τ' at $\{e_1, e_2\}$ exists and is unique as long as the tree S' obtained by folding S^τ along e_1 and e_2 has trivial edge stabilizers. As any carrying map factors through the folding map $g : S^\tau \to S'$, a sufficient condition for the existence of τ' is that τ carries a tree with trivial arc stabilizers. This covers all of the relevant train-tracks in this paper.

Definition 4.10. Given an admissible train-track τ, we denote by $\mathcal{P}(\tau)$ the subspace of ∂cv_N made of trees T with dense orbits such that τ is an ideal carrier of T.

In particular, trees in $\mathcal{P}(\tau)$ have trivial arc stabilizers.

Lemma 4.11. *Let τ be an admissible train-track, and let $e_1 = [v, v_1]$ and $e_2 = [v, v_2]$ be two edges that form an illegal turn in τ, such that $v_1 \sim^\tau v_2$, and such that* *the directions at v_1 and v_2 pointing towards v are τ-equivalent. If τ carries a tree with trivial arc stabilizers then there exists an admissible train-track τ' obtained from τ by a singular fold at $\{e_1, e_2\}$ such that $\mathcal{P}(\tau') = \mathcal{P}(\tau)$.*

Proof. As τ carries a tree with trivial arc stabilizers, following Remark 4.9 there exists a unique train track τ' obtained by a singular fold at $\{e_1, e_2\}$. Let $g : S^\tau \to S^{\tau'}$ be the folding map given in Definition 4.8. The numbers and the stabilizers of equivalence classes of vertices and gates are unchanged under g, so that $i(\tau') = i(\tau)$. To check admissibility of τ', let v' be a vertex in $S^{\tau'}$. Then there exists $v \in V(S^\tau)$ such that $v' = g(v)$, and the g-image of any tripod of legal axes at v is a tripod of legal axes at v' (because g sends legal turns to legal turns).

We finally show that $\mathcal{P}(\tau) = \mathcal{P}(\tau')$. Let $T \in \mathcal{P}(\tau)$, with carrying map $f : S^\tau \to T$. Since the extremities of e_1 and e_2 are equivalent, these two edges are mapped to the same segment in T. It follows that the carrying map f factors through the fold g to attain a map $f' : S^{\tau'} \to T$, and one can check that the train-track induced by f' is obtained from τ' by a finite (possibly trivial) sequence of specializations (using the same sequence of specializations as when passing from τ to τ_f). So T is carried by τ', and equality of the indices of τ and τ' shows that τ' is also an ideal carrier of T. It follows that $\mathcal{P}(\tau) \subseteq \mathcal{P}(\tau')$. Conversely, if τ' is an ideal carrier of T with carrying map f' then the composition $f = f' \circ g$ is such that τ_f is obtained from τ by a finite (possibly trivial) sequence of specializations, and it follows that $\mathcal{P}(\tau') \subseteq \mathcal{P}(\tau)$. □

Definition 4.12 (partial fold, see Figure 3). Let τ be a train-track, and let $\{e_1, e_2\}$ be an illegal turn at a vertex $v \in S^\tau$. A train-track τ' is obtained from τ by a *partial fold* at $\{e_1, e_2\}$ if

- there is a map $g : S^\tau \to S^{\tau'}$ that consists in equivariantly identifying a proper initial segment $[v, v_1']$ of e_1 with a proper initial segment $[v, v_2']$ of e_2, so that the vertex $v' = g(v_1') = g(v_2')$ is trivalent in $S^{\tau'}$ (all vertices of $S^{\tau'}$ that are not in the F_N-orbit of v' have a unique g-preimage in S^τ),
- for all $v, w \in V(S^\tau)$, we have $g(v) \sim_V^{\tau'} g(w)$ if and only if $v \sim_V^\tau w$,
- for all directions d, d' based at vertices of S^τ, we have $g(d') \sim^{\tau'} g(d)$ if and only if $d' \sim^\tau d$,
- the vertex v' is not equivalent to any other vertex in τ', and all directions at v' in $S^{\tau'}$ are pairwise inequivalent.

Remark 4.13. The last condition in the definition implies that the new vertex v' is exceptional in τ'. As with singular folds, such a partial fold τ' at $\{e_1, e_2\}$ exists and is unique as long as partially folding S^τ along e_1 and e_2 gives a tree with trivial arc stabilizers, which will be the case when $\mathcal{P}(\tau)$ is nonempty.

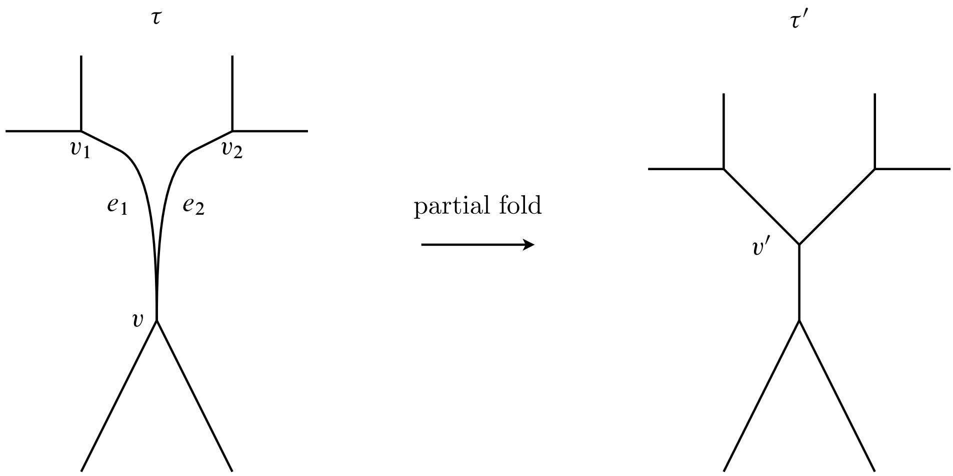


Figure 3. A partial fold.

Lemma 4.14. *Let τ be an admissible train-track, and let τ' be a train-track obtained from τ by a partial fold. Then τ' is admissible, and $i(\tau') = i(\tau)$.*

Proof. The condition from the definition of admissibility is easy to check at any vertex in $S^{\tau'}$ not in the orbit of v' (with the notations from Definition 4.12). Let d_0, d_1, d_2 be the three directions at v', containing the g-images of $v = v_0, v_1, v_2$, respectively. Let $i \in \{0, 1, 2\}$ (considered modulo 3). To construct a legal element whose axis crosses the turn $\{d_i, d_{i+1}\}$, we argue as in the proof of Lemma 2.9: one first finds two elements h, h' that are legal in both τ and τ' such that

- the axes $A_h, A_{h'}$ in $S^{\tau'}$ go through $g(v_i)$ and $g(v_{i+1})$ respectively,
- the axes A_h and $A_{h'}$ do not go through v',
- the subtree Y spanned by A_h and $A_{h'}$ (which contains the turn $\{d_i, d_{i+1}\}$) is legal.

The element hh' is then a legal element that crosses the turn $\{d_i, d_{i+1}\}$, as required.

To see that $i(\tau') = i(\tau)$, we first note that $\mathcal{A}(\tau') = \mathcal{A}(\tau)$ because the stabilizer of any equivalence class corresponding to a new vertex is trivial, and the stabilizers of the other equivalence classes of vertices are unchanged when passing from τ to τ'. In addition, under the folding process the number of nonexceptional equivalence classes of vertices, along with the ranks of their stabilizers and the number of associated gates, remains the same, and the new exceptional trivalent vertex contributes 0 to the index. Hence $i_{\text{geom}}(\tau') = i_{\text{geom}}(\tau)$, and therefore $i(\tau') = i(\tau)$ also. $\square$

Lemma 4.15. *Let τ be an admissible train-track, and let $T \in \mathcal{P}(\tau)$ with carrying map $f : S^\tau \to T$. Assume that there exist two edges e_1, e_2 in S^τ that form an illegal turn, such that $|f(e_1) \cap f(e_2)|$ is smaller than both $|f(e_1)|$ and $|f(e_2)|$. Then there* *exists a train-track τ' obtained from τ by applying a partial fold at $\{e_1, e_2\}$ such that $T \in \mathcal{P}(\tau')$.*

Proof. Recall that S^τ is equipped with the metric induced by the map $f : S^\tau \to T$, so that any edge e of S^τ has length $|f(e)|$. Let S' be the simplicial tree obtained from S^τ by equivariantly identifying a proper initial segment $[v, v_1']$ of e_1 of length $|f(e_1) \cap f(e_2)|$ with a proper initial segment $[v, v_2']$ of e_2 of the same length. Let $g : S^\tau \to S'$ be the folding map. We first claim that the vertex $v' := g(v_1') = g(v_2')$ is trivalent in S'. Indeed, otherwise, the vertex v' would have infinite stabilizer, so $\mathcal{A}(\tau)$ would not be a maximal free factor system elliptic in T, a contradiction. We thus have $S' = S^{\tau'}$. The map f factors through a map $f' : S^{\tau'} \to T$. Since $i(\tau') = i(\tau)$, it is enough to prove that the train-track $\tau_{f'}$ is obtained from τ' by a finite sequence of specializations.

By definition, the train-track τ_f is obtained from τ by a finite sequence of specializations. Let $(\tau_f)'$ be the partial fold of τ_f at $\{e_1, e_2\}$, and notice that $(\tau_f)'$ is obtained from τ' by the same sequence of specializations as when passing from τ to τ_f.

If the f'-image of the new trivalent vertex v' is not equal to the f-image of any other vertex in S^τ, then $\tau_{f'} = (\tau_f)'$ and we are done. Otherwise, we will show that $\tau_{f'}$ is a specialization of $(\tau_f)'$, from which the lemma will follow. Indeed, if $f'(v')$ is equal to the f-image of another vertex $v_0 \in S^\tau$, we first observe that we can always find such a v_0 which is not in the orbit of v'. This is because if v' is only identified with vertices in its own orbit, then the equivalence class of v' in $\tau_{f'}$ has nontrivial stabilizer in $\tau_{f'}$, and contributes positively to the geometric index of $\tau_{f'}$. Then $i(\tau_{f'}) > i(\tau)$, contradicting the fact that τ is an ideal carrier. So we can assume that v_0 and v' are not in the same orbit. Next, we observe that all germs of directions at v' are identified by f' with germs of directions at vertices that are τ-equivalent to v_0, as otherwise we would be creating a new gate when passing from τ to $\tau_{f'}$, again increasing the index. So $\tau_{f'}$ is a specialization of $(\tau_f)'$. □

Definition 4.16 (full fold, see Figure 4). Let τ be a train-track, and let $\{e_1, e_2\}$ be an illegal turn at a vertex $v \in S^\tau$. Denote by v_1, v_2 the other extremities of e_1, e_2. Let d_0 be a direction in τ based at a vertex τ-equivalent to v_1. A train-track τ' is obtained from τ by a *full fold of* e_1 *into* e_2 *with special gate* $[d_0]$ if

- there is an F_N-equivariant map $g : S^\tau \to S^{\tau'}$ that consists in equivariantly identifying $[v, v_1]$ with a proper initial segment $[v, v_2'] \subsetneq [v, v_2]$,
- for all vertices $v, v' \in V(S^\tau)$, we have $g(v') \sim^{\tau'} g(v)$ if and only if $v' \sim^\tau v$,
- for all directions d, d' based at vertices of S^τ, we have $g(d') \sim^{\tau'} g(d)$ if and only if $d' \sim^\tau d$,
- if d is the direction in $S^{\tau'}$ based at $g(v_1) = g(v_2')$ and pointing towards $g(v_2)$, then $d \sim g(d_0)$.

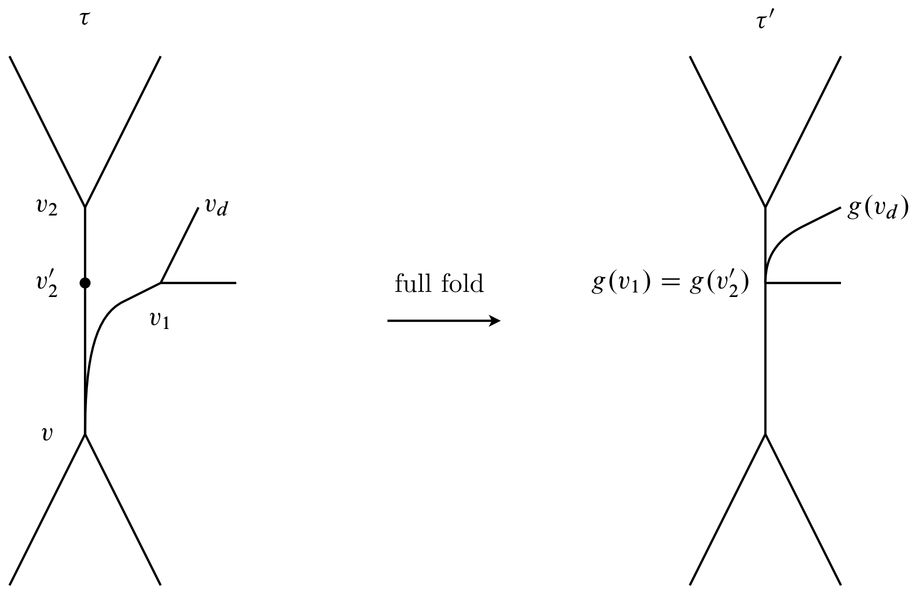


Figure 4. A full fold. The new direction at $g(v_1) = g(v_2')$ has been identified with the g-image of the direction d_0 at v_1 with initial edge $[v_1, v_d]$ (in general, the direction d_0 may also be based at a vertex $v_1' \neq v_1$ as long as v_1 and v_1' are τ-equivalent).

The number of equivalence classes of vertices remains unchanged under a full fold, and the vertex stabilizers in $S^{\tau'}$ are the same as the vertex stabilizers in S^τ. The fold creates no new gates, so the index remains unchanged. If τ is admissible then τ' is also admissible. Hence:

Lemma 4.17. *Let τ be an admissible train-track, and let τ' be a train-track obtained from τ by a full fold. Then τ' is admissible and $i(\tau') = i(\tau)$.* □

Lemma 4.18. *Let τ be an admissible train-track, and let $T \in \mathcal{P}(\tau)$ with carrying map $f : S^\tau \to T$. Let $e_1 = [v, v_1]$ and $e_2 = [v, v_2]$ be two edges in S^τ that form an illegal turn. Assume that $f(e_1) \subsetneq f(e_2)$. Also assume that there is a direction d_0 (whose initial edge we denote by $[v_1', v_d]$) at a vertex $v_1' \sim^\tau v_1$ such that*

$$d_T(f(v_d), f(v_2)) < d_T(f(v_d), f(v_1')) + d_T(f(v_1'), f(v_2)). \tag{7}$$

Then there exists a train-track τ' obtained from τ by fully folding e_1 into e_2, with special gate $[d_0]$ such that $T \in \mathcal{P}(\tau')$.

Proof. As in the previous cases, existence of τ' follows from the fact that τ carries a tree with trivial arc stabilizers. Since $f(e_1) \subsetneq f(e_2)$, the map f factors through the fold $g : S^\tau \to S^{\tau'}$ to reach an F_N-equivariant map $f' : S^{\tau'} \to T$, and (7) ensures that f' identifies germs of $[g(v_1), g(v_2)]$ and of $g(d_0)$. So $\tau_{f'}$ is obtained from τ' by the same finite sequence of specializations as when passing from τ to τ_f. We know in addition that $i(\tau') = i(\tau)$ (Lemma 4.17). This implies that $T \in \mathcal{P}(\tau')$. $\square$

Corollary 4.19. *Let τ be an admissible train-track, and let $T \in \mathcal{P}(\tau)$, with carrying map $f : S^\tau \to T$. Let $e_1 = [v, v_1]$ and $e_2 = [v, v_2]$ be two edges in S^τ that form an illegal turn. Assume that $f(e_1) \subsetneq f(e_2)$. Also assume that T is not carried by any specialization of τ. Then there exists a train-track τ' obtained from τ by fully folding e_1 into e_2, such that $T \in \mathcal{P}(\tau')$.*

Proof. Let S' be the underlying simplicial tree of any full fold of τ. Since $f(e_1) \subsetneq f(e_2)$, the map f factors through an F_N-equivariant map $f' : S^{\tau'} \to T$. Equality in indices shows the existence of a direction d_0 in τ for which (7) holds, as otherwise the fold would create a new gate, and T would be carried by a train-track with higher index. The direction d_0 is based at a vertex v_1' satisfying $f(v_1') = f(v_1)$. Since T is not carried by any specialization of τ, this implies that v_1' is τ-equivalent to v_1. Therefore, Lemma 4.18 implies that $T \in \mathcal{P}(\tau')$, where τ' is the full fold of τ with special gate $[d_0]$. $\square$

4A3. *Folding sequences: Proof of Lemma 4.3.* Given two train-tracks τ and τ', we say that τ' is obtained from τ by *applying a folding move* if τ' is obtained from τ by applying either a singular fold, a partial fold or a full fold. We then define a *folding sequence of train-tracks* as in Definition 4.2: recall that this is a sequence $(\tau_i)_{i \in \mathbb{N}}$ of train-tracks such that τ_{i+1} is obtained from τ_i by applying a folding move, followed by a finite (possibly trivial) sequence of specializations. The goal of the present section is to prove Lemma 4.3. We will actually prove a slightly stronger version of it, given in Lemma 4.22. The main fact we will use in the proof is the following.

Lemma 4.20. *Let τ be an admissible train-track, and let $T \in \mathcal{P}(\tau)$. Then there exists a train-track $\tau' \neq \tau$ obtained either by a specialization, or by applying a folding move, such that $T \in \mathcal{P}(\tau')$.*

Proof. Assume that T is not carried by any train-track $\tau' \neq \tau$ obtained from τ by performing a specialization. Let $f : S^\tau \to T$ be the carrying map, and let $e_1 = [v, v_1]$ and $e_2 = [v, v_2]$ be two edges that form an illegal turn in τ: this exists because T is not simplicial by definition of $\mathcal{P}(\tau)$. If $f(v_1) = f(v_2)$, then $v_1 \sim^\tau v_2$, as otherwise T would be carried by a specialization of τ. In this case f identifies germs of the directions at v_1 and v_2 pointing towards v, so Lemma 4.11 implies that $T \in \mathcal{P}(\tau')$, where τ' is the singular fold of τ. If $f(v_1) \neq f(v_2)$, and if $|f(e_1) \cap f(e_2)|$ is smaller than both $|f(e_1)|$ and $|f(e_2)|$, then Lemma 4.15 implies that T is carried by the partial fold of τ at $\{e_1, e_2\}$. Finally, if $f(e_1) \subsetneq f(e_2)$ (or vice versa), then Corollary 4.19 implies that T is carried by a full fold of τ. $\square$

We also make the following observation.

Lemma 4.21. *Let τ be an admissible train-track, and let τ' be a train-track obtained from τ by applying either a specialization or a folding move. Then $\mathcal{P}(\tau') \subseteq \mathcal{P}(\tau)$ (and hence $P(\tau') \subseteq P(\tau)$).*

Proof. If τ' is obtained from τ by applying a specialization, then the conclusion follows from the definition of being carried, together with the fact that $i(\tau') = i(\tau)$ (Lemma 4.7). If τ' is obtained from τ by applying a folding move, then the conclusion follows from the fact that $i(\tau') = i(\tau)$ (Lemmas 4.11, 4.14 and 4.17), together with the observation that if $g : S^\tau \to S^{\tau'}$ is the fold map, and $f : S^{\tau'} \to T$ is a carrying map, then $f \circ g : S^\tau \to T$ is also a carrying map. □

We are now in position to prove the following stronger version of Lemma 4.3.

Lemma 4.22. *Let $\xi \in \partial_\infty FZ_N$. Let $\tau_0, \dots, \tau_k$ be a finite sequence of admissible train-tracks such that $\xi \in P(\tau_k)$, and for each $i \in \{0, \dots, k-1\}$, the train-track τ_{i+1} is obtained from τ_i by applying a folding move followed by a finite (possibly trivial) sequence of specializations. Then there exists a folding sequence $(\tau'_i)_{i \in \mathbb{N}}$ of train-tracks directed by ξ, such that $\tau_i = \tau'_i$ for all $i \in \{0, \dots, k-1\}$, and τ'_k is obtained from τ_k by a finite (possibly trivial) sequence of specializations.*

Proof. Lemma 4.21 shows that $\xi \in P(\tau_i)$ for all $i \in \{1, \dots, k\}$. The conclusion is then obtained by iteratively applying Lemma 4.20 to some mixing representative of the class ξ (starting with the train-track τ_k), and noticing that we can only perform finitely many specializations in a row. □

4B. *Openness of the set of points carried by a specialization or a fold.*

4B1. *From trees to equivalence classes of trees.* We first reduce the proofs of Lemmas 4.4 and 4.5 to their analogous versions for trees in $\overline{cv}_N$.

Lemma 4.23. *Let τ and τ' be admissible train-tracks. If $\mathcal{P}(\tau')$ is an open subset of $\mathcal{P}(\tau)$ in ∂cv_N, then $P(\tau')$ is an open subset of $P(\tau)$ in $\mathcal{X}_N/{\approx}$.*

Proof. The conclusion is obvious if $\mathcal{P}(\tau')$ is empty (in this case $P(\tau')$ is also empty), so we assume otherwise. Since $\mathcal{P}(\tau') \subseteq \mathcal{P}(\tau)$, we have $i(\tau') = i(\tau)$. Let $\xi \in P(\tau')$, and let $(\xi_n)_{n \in \mathbb{N}} \in P(\tau)^{\mathbb{N}}$ be a sequence converging to ξ. We wish to prove that $\xi_n \in P(\tau')$ for all sufficiently large $n \in \mathbb{N}$ (we may use sequential arguments as $\mathcal{X}_N/{\approx}$ is a separable metric space).

For all $n \in \mathbb{N}$, let T_n be a mixing representative of ξ_n. Since $\overline{cv}_N$ is projectively compact, up to passing to a subsequence, we can assume that $(T_n)_{n \in \mathbb{N}}$ converges projectively to a tree $T \in \overline{cv}_N$. Closedness of the boundary map $\partial\pi : \mathcal{X}_N \to \partial_\infty FZ_N$ shows that $T \in \mathcal{X}_N$, and T is a representative of the class ξ. Since $\mathcal{A}(\tau)$ is elliptic in each tree T_n, it is also elliptic in T. If T is not mixing, then Corollary 1.9 implies that $\mathcal{A}(\tau)$ is not the largest free factor system elliptic in the mixing representatives of ξ, contradicting that $i(\xi) = i(\tau)$. So T is mixing, and the fact that $\xi \in P(\tau')$ thus implies that $T \in \mathcal{P}(\tau')$. Openness of $\mathcal{P}(\tau')$ in $\mathcal{P}(\tau)$ then shows that $T_n \in \mathcal{P}(\tau')$ for all sufficiently large $n \in \mathbb{N}$. Since T_n is a mixing representative of ξ_n, this precisely means that $\xi_n \in P(\tau')$ for all sufficiently large $n \in \mathbb{N}$, as desired. □

4B2. *Specializations.* In this section, we will prove Lemma 4.4. More precisely, we will prove its analogous version for trees in $\overline{cv}_N$, from which Lemma 4.4 follows thanks to Lemma 4.23.

Lemma 4.24. *Let τ be an admissible train-track, and let τ' be a specialization of τ. Then $\mathcal{P}(\tau')$ is an open subset of $\mathcal{P}(\tau)$.*

Proof. By Lemma 4.21, we have $\mathcal{P}(\tau') \subseteq \mathcal{P}(\tau)$. Let $T \in \mathcal{P}(\tau')$, and let $(T_n)_{n\in\mathbb{N}} \in \mathcal{P}(\tau)^{\mathbb{N}}$ be a sequence of trees that converges to T. We aim to show that T_n is carried by τ' for all sufficiently large $n \in \mathbb{N}$, which will imply that $T_n \in \mathcal{P}(\tau')$ by equality of the indices. Let $[v_0]$ be the class of vertices in S^τ at which the specialization occurs, and let $v_1 \in V(S^\tau)$ be a vertex identified with v_0 in τ'.

Let e_0^1, e_0^2, e_0^3 be three edges adjacent to vertices in the class $[v_0]$, determining distinct gates, and let e_1^1, e_1^2, e_1^3 be edges based at vertices τ-equivalent to v_1 that are identified with e_0^1, e_0^2, e_0^3 in τ'. We aim to show that for all sufficiently large $n \in \mathbb{N}$, the vertices v_0 and v_1 are identified by the carrying map $f_n : S^\tau \to T_n$, and so are initial segments of e_0^i and e_1^i for all $i \in \{1, 2, 3\}$.

For all $i \in \{1, 2, 3\}$, the intersection $f(e_0^i) \cap f(e_1^i)$ (where $f : S^\tau \to T$ is the carrying map) is nondegenerate. Since the carrying map varies continuously with the carried tree (Lemma 2.11), for all sufficiently large $n \in \mathbb{N}$, the image $f_n(e_1^i)$ has nondegenerate intersection with $f_n(e_0^i)$. Since the three directions determined by e_0^1, e_0^2, e_0^3 are inequivalent, their images $f_n(e_0^1), f_n(e_0^2), f_n(e_0^3)$ form a tripod at $f_n(v_0)$. If $f_n(v_1) \neq f_n(v_0)$, then at least two of the images $f_n(e_1^1), f_n(e_1^2), f_n(e_1^3)$ would then have a nondegenerate intersection (containing $[f_n(v_0), f_n(v_1)]$), which is a contradiction because these three edges are inequivalent. So $f_n(v_1) = f_n(v_0)$, and since $f_n(e_1^i)$ has nondegenerate intersection with $f_n(e_0^i)$, it follows that T_n is carried by τ'. □

4B3. *Folds.* Lemma 4.5 follows from the following proposition together with Lemma 4.23.

Proposition 4.25. *Let τ be an admissible train-track. Let τ' be a train-track obtained from τ by folding an illegal turn. Then $\mathcal{P}(\tau')$ is open in $\mathcal{P}(\tau)$.*

Proof. By Lemma 4.21, we have $\mathcal{P}(\tau') \subseteq \mathcal{P}(\tau)$. If τ' is obtained from τ by applying a singular fold, then Lemma 4.11 ensures that $\mathcal{P}(\tau') = \mathcal{P}(\tau)$, so the conclusion holds. Denoting by $f : S^\tau \to T$ the carrying map, we can thus assume that $f(v_1) \neq f(v_2)$ (using the notations of the previous sections). Let $p := f(e_1) \cap f(e_2)$. The endpoints of p, $f(e_1)$, and $f(e_2)$ are branch points in T (because τ is admissible), so the lengths of these segments are determined by a finite set of translation lengths in T.

If τ' is obtained from τ by a partial fold, then $|p|$ is strictly less than both $|f(e_1)|$ and $|f(e_2)|$. Using the fact that the carrying map f varies continuously with the carried tree (Lemma 2.11), we see that this property remains true for all trees in a neighborhood of T, and Lemma 4.15 implies that all trees in this neighborhood are carried by τ'.

Assume now that τ' is obtained from τ by a full fold, with e_1 fully folded into e_2. Let v_d be the endpoint of an edge corresponding to a direction d based at a vertex v_1' equivalent to v_1, in the special gate. Then $|f(e_1)| < |f(e_2)|$, and $d_T(f(v_d), f(v_2)) < d_T(f(v_d), f(v_1')) + d_T(f(v_1'), f(v_2))$. These are open conditions in $\overline{cv}_N$. Therefore, Lemma 4.18 gives the existence of an open neighborhood U of T in $\mathcal{P}(\tau)$ such that all trees in U are carried by the same full fold. □

4C. ***Control on the diameter: getting finer and finer decompositions.*** The goal of the present section is to prove Lemma 4.6. The key lemma is the following.

Lemma 4.26. *Let $\xi \in \partial_\infty FZ_N$, and let $(\tau_i)_{i\in\mathbb{N}}$ be a folding sequence of train-tracks directed by ξ. Then (S^{τ_i}) converges to ξ for the topology on $FZ_N \cup \partial_\infty FZ_N$.*

Proof. Let T be a mixing representative of ξ. We can find a sequence $(S_n)_{n\in\mathbb{N}}$ of simplicial metric F_N-trees with trivial edge stabilizers such that

- for all $n \in \mathbb{N}$, the simplicial tree S^{τ_n} is obtained from S_n by forgetting the metric,
- for all $n \in \mathbb{N}$, the unique carrying map $f_n : S_n \to T$ is isometric on edges.

In addition, for all $i < j$, there are natural morphisms $f_{ij} : S_i \to S_j$, such that $f_{ik} = f_{jk} \circ f_{ij}$ for all $i < j < k$. The sequence $(S_n)_{n\in\mathbb{N}}$ converges to a tree S_∞: indeed, for all $g \in F_N$, the sequence $(\|g\|_{S_n})_{n\in\mathbb{N}}$ is nonincreasing, so it converges. In addition, the tree S_∞ is not reduced to a point, because the legal structure on S_0 induced by the morphisms f_{0j} will stabilize as j goes to $+\infty$, and a legal element (which exists because τ_0 is admissible, and every legal turn for the train-track τ_0 is also legal for the train-track structure on S_0 induced by the morphisms f_{0j}) cannot become elliptic in the limit. By taking a limit of the maps f_{0j}, we get an F_N-equivariant map $f_{0\infty} : S_0 \to S_\infty$, which is isometric when restricted to every edge of S_0.

We will show that S_∞ is $\mathcal{Z}$-averse and $\approx$-equivalent to T, which implies the convergence of S^{τ_i} to ξ (for the topology on $FZ_N \cup \partial_\infty FZ_N$) by the continuity statement for the boundary map $\partial\pi$ given in Theorem 1.3.

We first claim that the tree S_∞ is not simplicial. Indeed, assume towards a contradiction that it is, and let S'_∞ be the simplicial tree obtained from S_∞ by adding the $f_{0\infty}$-images of all vertices of S_0 to subdivide the edges of S_∞. Subdivide the edges of S_0 so that every edge of S_0 is mapped to an edge of S'_∞. Since T has trivial arc stabilizers, the folding process from S_0 to S'_∞ never identifies two edges in the same orbit, so there is a lower bound on the difference between the volume of S_{i+1}/F_N

and the volume of S_i/F_N. Therefore, the folding process has to stop, contradicting the fact that the folding sequence is infinite. This shows that S_∞ is not simplicial.

We now assume towards a contradiction that S_∞ is not equivalent to T. There exists a 1-Lipschitz F_N-equivariant map f from S_∞ to the metric completion of T, obtained by taking a limit of the maps $f_{n\infty}$ as n goes to infinity; see the construction from [Horbez 2016b, Theorem 4.3]. Therefore, if S_∞ has dense orbits, then there is a 1-Lipschitz F_N-equivariant alignment-preserving map from S_∞ to T by [Horbez 2016b, Proposition 5.7], which implies that S_∞ is $\mathcal{Z}$-averse and equivalent to T by [Horbez 2016a, Theorem 4.1]. Therefore, its remains to show that S_∞ has dense orbits. Assume towards a contradiction that it does not. Since S_∞ is a limit of free and simplicial F_N-trees that all admit 1-Lipschitz maps onto it, it has trivial arc stabilizers. Since S_∞ is not simplicial, there is a free factor A acting with dense orbits on its minimal subtree S_A in S_∞. By [Reynolds 2012, Lemma 3.10], the translates in $\overline{T}$ of $f(S_A)$ (which cannot be reduced to a point) form a transverse family in $\overline{T}$: this is a contradiction because the stabilizer of a subtree in a transverse family in a mixing $\mathcal{Z}$-averse tree cannot be a free factor (in fact, it cannot be elliptic in any $\mathcal{Z}$-splitting of F_N by [Horbez 2016a, Proposition 4.23]). □

Proof of Lemma 4.6. Let $\xi' \in P(\tau_i)$, and let T' be a mixing representative of ξ'. By Lemma 4.22, there exists a folding sequence of train-tracks $(\tau'_j)_{j\in\mathbb{N}}$ directed by ξ', such that $\tau'_j = \tau_j$ for all $j < i$, and τ'_i is obtained from τ_i by a finite (possibly trivial) sequence of specializations. All trees $S^{\tau'_i}$ then lie on the image in FZ_N of an optimal liberal folding path, which is a quasigeodesic by [Mann 2014a], and $(S^{\tau'_i})_{i\in\mathbb{N}}$ converges to ξ' by Lemma 4.26. This shows that any quasigeodesic ray in FZ_N from S^{τ_0} to a point in $P(\tau_i)$ passes within bounded distance of S^{τ_i}. As S^{τ_i} converges to $\xi \in \partial_\infty FZ_N$ (Lemma 4.26), it follows from the definition of a visual distance that the diameter of $P(\tau_i)$ converges to 0. □

5. End of the proof of the main theorem

We now sum up the arguments from the previous sections to complete the proof of our main theorem.

Proof of the main theorem. We start with the case of FZ_N. The Gromov boundary $\partial_\infty FZ_N$ is a separable metric space (equipped with a visual metric). It can be written as the union of all strata X_i (made of boundary points of index i), where $i \in \mathbb{N}^2$ can only take finitely many values. Each stratum X_i is the union of the sets $P(\tau)$, where τ varies over the collection of all train-tracks of index i, and the collection of all train-tracks τ for which $P(\tau)$ is nonempty is countable by Lemma 2.8. By Proposition 3.1, each cell $P(\tau)$ is closed in its stratum X_i, and by Proposition 4.1 it has dimension at most 0. So each X_i is a countable union of closed 0-dimensional subsets, so X_i is 0-dimensional by the countable union theorem [Engelking 1978, Lemma 1.5.2]. Since $\partial_\infty FZ_N$ is a union of finitely many subsets of dimension 0, the union theorem [Engelking 1978, Proposition 1.5.3] implies that $\partial_\infty FZ_N$ has finite topological dimension (bounded by the number of strata, minus 1). In particular, the topological dimension of $\partial_\infty FZ_N$ is equal to its cohomological dimension; see the discussion in [Engelking 1978, pp. 94–95]. This gives the desired bound because the cohomological dimension of $\partial_\infty FZ_N$ is bounded by $3N-5$ by [Horbez 2016a, Corollary 7.3] (this is proved by using the existence of a cell-like map from a subset of ∂CV_N — whose topological dimension is equal to $3N-5$ by [Gaboriau and Levitt 1995], and applying [Rubin and Schapiro 1987]).

It was shown in Section 2B3 that the Gromov boundary $\partial_\infty FF_N$ is equal to the union of the strata X_i, with i comprised between $(0,0)$ and $(0,2N-2)$. Therefore, the above argument directly shows that the topological dimension of $\partial_\infty FF_N$ is at most $2N-2$ (without appealing to cohomological dimension).

Finally, the Gromov boundary $\partial_\infty I_N$ is equal to a union of strata X'_i, with $i \in \mathbb{Z}_+$ comprised between 0 and $2N-3$, where each stratum X'_i is a subspace of the stratum $X_{(0,i)}$ from above. Given a train-track τ of index $(0,i)$, we let $P'(\tau) := P(\tau) \cap X'_i$. Being a subspace of $P(\tau)$, the set $P'(\tau)$ has dimension at most 0. In addition $P'(\tau)$ is closed in its stratum X'_i, because its boundary in $\partial_\infty I_N$ is made of points of higher index by Proposition 3.1. The same argument as above thus shows that the topological dimension of $\partial_\infty I_N$ is at most $2N-3$. □

Appendix: Why equivalence classes of vertices and specializations?

In this appendix, we would like to illustrate the reason why we needed to introduce an equivalence relation on the vertex set of S^τ in the definition of a train-track (Definition 2.1), and allow for specializations in the definition of carrying (Definition 2.6).

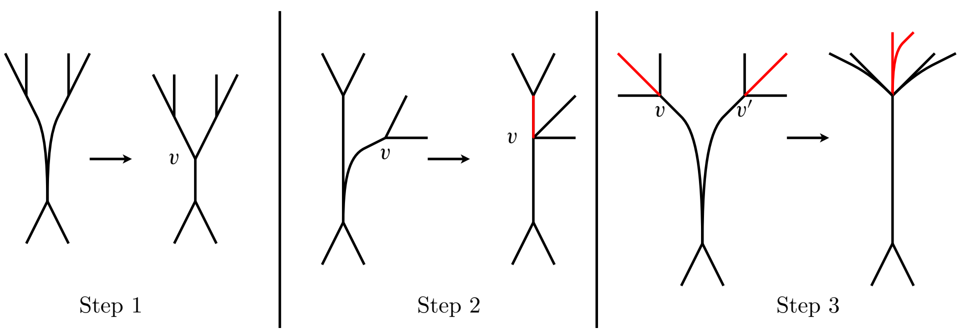


Figure 5. The following three steps may repeat infinitely often along a folding sequence, showing that the homeomorphism type of the underlying simplicial tree might never stabilize.

Why equivalence classes of vertices in the definition of a train-track? In the world of train-tracks on surfaces, when one defines a train-track splitting sequence towards an arational foliation, the homeomorphism type of the complement of the train-track eventually stabilizes, after which each singularity in the limiting foliation is determined by a complementary region of a track, and the visible prongs determine the index of the singular point.

On the contrary, in a folding sequence of train-tracks towards an arational tree T as defined in the present paper, the number of vertices in the preimage of a branch point in T, as well as the number of directions at these vertices, may never stabilize, as illustrated by the following situation (see Figure 5).

At some point on the folding sequence, one may have to perform a singular fold as depicted on the right-hand side of Figure 5 (Step 3). If we had not declared the two identified vertices equivalent before computing the index, such an operation would have resulted in a drop in index. If such a situation could only occur finitely many times along the folding sequence, we could have defined a "stable" index along the folding sequence, but this might not be true in general. Indeed, later on along the folding sequence, a new trivalent vertex v can be created due to a partial fold (Step 1 on Figure 5), and this exceptional vertex v may then be declared equivalent to another vertex v' in the track by applying a specialization: this results in the possibility of performing a new singular fold later on, identifying v and v'. One could then hope that eventually, all singular folds in the folding sequence involve an exceptional vertex, which would not cause trouble as far as index is concerned; but even this may fail to be true in general. Indeed, it might happen that later on in the process, a full fold involving v creates a fourth direction at v (this is the red direction in Step 2 of Figure 5), which is not equivalent to any direction at v (but is instead equivalent to a direction at another vertex v' in the same class as v).

Introducing an equivalence class on the set of vertices in the definition of a train-track ensures that the index of our train-tracks remains constant along a folding path, and prevents overcounting the number of branch points in the limiting tree when counting branch points in the train-track.

Why introduce specializations in the definition of carrying? We would finally like to give a word of motivation for the necessity to introduce specializations in the definition of carrying, instead of just saying that a train-track τ carries a tree T if $\tau = \tau_f$ for some F_N-equivariant map $f : S^\tau \to T$. As already observed, in a folding sequence of train-tracks directed by T, partial folds create new trivalent vertices, and it can happen that the newly created vertex v gets mapped to the same point in T as another vertex $v' \in S^\tau$. We could have tried in this case to perform the partial fold and the specialization at the same time, in other words declare that there are several distinct partial folds of τ, including one (denoted by τ_1) where v is not declared equivalent with any other vertex, and a second (denoted by τ_2) where v is declared equivalent to v' (and there would be infinitely many other partial folds where v is identified with any possible vertex from S^τ). With this approach, the difficulty comes when proving openness of the set $\mathcal{P}(\tau_1)$ within $\mathcal{P}(\tau)$. Indeed, for a certain point $T \in \mathcal{P}(\tau)$, the carrying map may not identify the new trivalent vertex v in $S^{\tau'}$ with any other vertex in $S^{\tau'}$ (hence T is carried by τ_1), while for nearby trees T_n in $\mathcal{P}(\tau)$ (for the Gromov–Hausdorff topology), the carrying map identifies v with a vertex v_n going further and further away in S^τ as n goes to infinity (hence no tree T_n is carried by τ_1). So without the extra flexibility in the definition of carrying, this would lead to $\mathcal{P}(\tau_1)$ not being open in $\mathcal{P}(\tau)$. This justifies our definition of carrying. We have a single way of performing a partial fold of a turn, and the specialization is performed later on along the folding sequence if needed.

Acknowledgments

We would like to thank Patrick Reynolds for conversations we had related to the present project, and Vera Tonić for pointing out to us the reference [Engelking 1978] relating topological and cohomological dimensions. The present work was supported by the National Science Foundation under Grant No. DMS-1440140 while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2016 semester. Bestvina was supported by the NSF under Grant No. DMS-1607236.

References

[Bell and Fujiwara 2008] G. C. Bell and K. Fujiwara, "The asymptotic dimension of a curve graph is finite", *J. Lond. Math. Soc.* (2) **77**:1 (2008), 33–50. MR

[Bestvina and Bromberg 2019] M. Bestvina and K. Bromberg, "On the asymptotic dimension of the curve complex", *Geom. Topol.* **23**:5 (2019), 2227–2276. MR Zbl

[Bestvina and Feighn 1992] M. Bestvina and M. Feighn, "Outer limits", preprint, 1992, available at https://tinyurl.com/bestouter.

[Bestvina and Feighn 2014] M. Bestvina and M. Feighn, "Hyperbolicity of the complex of free factors", *Adv. Math.* **256** (2014), 104–155. MR Zbl

[Bestvina and Reynolds 2015] M. Bestvina and P. Reynolds, "The boundary of the complex of free factors", *Duke Math. J.* **164**:11 (2015), 2213–2251. MR Zbl

[Bestvina et al. 2015] M. Bestvina, K. Bromberg, and K. Fujiwara, "Constructing group actions on quasi-trees and applications to mapping class groups", *Publ. Math. Inst. Hautes Études Sci.* **122** (2015), 1–64. MR Zbl

[Bestvina et al. 2017] M. Bestvina, V. Guirardel, and C. Horbez, "Boundary amenability of Out(F_N)", preprint, 2017. arXiv

[Bestvina et al. 2019] M. Bestvina, K. Bromberg, and K. Fujiwara, "Proper actions on finite products of quasi-trees", preprint, 2019. arXiv

[Cohen and Lustig 1995] M. M. Cohen and M. Lustig, "Very small group actions on $\mathbb{R}$-trees and Dehn twist automorphisms", *Topology* **34**:3 (1995), 575–617. MR Zbl

[Coulbois and Hilion 2012] T. Coulbois and A. Hilion, "Botany of irreducible automorphisms of free groups", *Pacific J. Math.* **256**:2 (2012), 291–307. MR Zbl

[Culler and Morgan 1987] M. Culler and J. W. Morgan, "Group actions on ℝ-trees", *Proc. Lond. Math. Soc.* (3) **55**:3 (1987), 571–604. MR Zbl

[Dowdall and Taylor 2017] S. Dowdall and S. J. Taylor, "The co-surface graph and the geometry of hyperbolic free group extensions", *J. Topol.* **10**:2 (2017), 447–482. MR Zbl

[Engelking 1978] R. Engelking, *Dimension theory*, North-Holland Math. Library **19**, North-Holland, Amsterdam, 1978. MR Zbl

[Gabai 2009] D. Gabai, "Almost filling laminations and the connectivity of ending lamination space", *Geom. Topol.* **13**:2 (2009), 1017–1041. MR Zbl

[Gaboriau and Levitt 1995] D. Gaboriau and G. Levitt, "The rank of actions on ℝ-trees", *Ann. Sci. École Norm. Sup.* (4) **28**:5 (1995), 549–570. MR Zbl

[Guirardel and Horbez 2019] V. Guirardel and C. Horbez, "Boundaries of relative factor graphs and subgroup classification for automorphisms of free products", preprint, 2019. arXiv

[Hamenstädt 2012] U. Hamenstädt, "The boundary of the free factor graph and the free splitting graph", preprint, 2012. arXiv

[Hamenstädt 2019] U. Hamenstädt, "Asymptotic dimension and the disk graph, II", *J. Topol.* **12**:3 (2019), 675–684. MR Zbl

[Horbez 2016a] C. Horbez, "Hyperbolic graphs for free products, and the Gromov boundary of the graph of cyclic splittings", *J. Topol.* **9**:2 (2016), 401–450. MR Zbl

[Horbez 2016b] C. Horbez, "Spectral rigidity for primitive elements of F_N", *J. Group Theory* **19**:1 (2016), 55–123. MR Zbl

[Horbez 2017] C. Horbez, "The boundary of the outer space of a free product", *Israel J. Math.* **221**:1 (2017), 179–234. MR Zbl

[Kapovich and Lustig 2009] I. Kapovich and M. Lustig, "Geometric intersection number and analogues of the curve complex for free groups", *Geom. Topol.* **13**:3 (2009), 1805–1833. MR Zbl

[Klarreich 1999] E. Klarreich, "The boundary at infinity of the curve complex and the relative Teichmüller space", preprint, 1999. To appear in *Groups Geom. Dyn.* arXiv

[Leininger and Schleimer 2009] C. J. Leininger and S. Schleimer, "Connectivity of the space of ending laminations", *Duke Math. J.* **150**:3 (2009), 533–575. MR Zbl

[Levitt 1994] G. Levitt, "Graphs of actions on ℝ-trees", *Comment. Math. Helv.* **69**:1 (1994), 28–38. MR Zbl

[Mann 2014a] B. Mann, "Hyperbolicity of the cyclic splitting graph", *Geom. Dedicata* **173** (2014), 271–280. MR Zbl

[Mann 2014b] B. Mann, *Some hyperbolic* Out(F_N)*-graphs and nonunique ergodicity of very small* F_N*-trees*, Ph.D. thesis, University of Utah, 2014, available at https://search.proquest.com/docview/1614969728.

[Masur and Minsky 1999] H. A. Masur and Y. N. Minsky, "Geometry of the complex of curves, I: Hyperbolicity", *Invent. Math.* **138**:1 (1999), 103–149. MR Zbl

[Paulin 1989] F. Paulin, "The Gromov topology on ℝ-trees", *Topology Appl.* **32**:3 (1989), 197–221. MR Zbl

[Reynolds 2011] P. Reynolds, "On indecomposable trees in the boundary of outer space", *Geom. Dedicata* **153** (2011), 59–71. MR Zbl

[Reynolds 2012] P. Reynolds, "Reducing systems for very small trees", preprint, 2012. arXiv

[Rubin and Schapiro 1987] L. R. Rubin and P. J. Schapiro, "Cell-like maps onto noncompact spaces of finite cohomological dimension", *Topology Appl.* **27**:3 (1987), 221–244. MR Zbl

Received February 12, 2019. Revised December 20, 2019.

MLADEN BESTVINA
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UT
UNITED STATES
bestvina@math.utah.edu

CAMILLE HORBEZ
LABORATOIRE DE MATHÉMATIQUES D'ORSAY
UNIVERSITÉ PARIS SUD, CNRS
UNIVERSITÉ PARIS-SACLAY
ORSAY
FRANCE
camille.horbez@math.u-psud.fr

RICHARD D. WADE
MATHEMATICAL INSTITUTE
UNIVERSITY OF OXFORD
OXFORD
UNITED KINGDOM
wade@maths.ox.ac.uk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 308, No. 1, 2020
https://doi.org/10.2140/pjm.2020.308.41

ON THE FIXED LOCUS OF FRAMED INSTANTON SHEAVES ON $\mathbb{P}^3$

ABDELMOUBINE AMAR HENNI

Let $\mathbb{T}$ be the three-dimensional torus acting on $\mathbb{P}^3$ and $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c)$ be the fixed locus of the corresponding action on the moduli space of rank 2 framed instanton sheaves on $\mathbb{P}^3$. We prove that $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c)$ consist only of non-locally-free instanton sheaves whose double dual is the trivial bundle $\mathcal{O}^{\oplus 2}_{\mathbb{P}^3}$. Moreover, we relate these instantons to Pandharipande–Thomas stable pairs and give a classification of their support. This allows us to compute a lower bound on the number of components of $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c)$.

1. Introduction

Framed instanton sheaves have been the subject of study for more than four decades and by many authors of different backgrounds. One of the main reasons is that they reflect a deep connection between algebraic geometry and mathematical physics; Atiyah, Drinfeld, Hitchin and Manin [Atiyah et al. 1978] fully classified the Yang–Mills anti-self-dual solutions, known as *instantons* (see also [Belavin et al. 1975; Donaldson and Kronheimer 1990]), on the four sphere S^4. The classification was given, first, by relating instantons with certain holomorphic bundles on the projective space $\mathbb{P}^3$, over $\mathbb{C}$, by means of Penrose–Ward correspondence. Then by using Horrocks *monads* [1964] the authors got linear algebraic data, called the *ADHM* data. Donaldson [1984] discovered that *framed* instantons on the four sphere S^4 correspond to some *framed* holomorphic bundles on the projective plane $\mathbb{P}^2$. Moreover, Nakajima [1994; 1999] considered framed sheaves in order to provide a

MSC2010: 14F05, 14J10.
Keywords: instantons, fixed locus, moduli spaces.

compactification of the moduli space of framed instanton bundles on surfaces. This led to the computation of many invariants [Nakajima 1999; Nakajima and Yoshioka 2005], such as Betti numbers and Euler characteristic of these moduli spaces, on one hand, and a connection to representation theory by means of quiver varieties [Nakajima 2011] and the infinite Heisenberg algebra [Nakajima 1999; Baranovsky 2000], on the other hand. It is worth to mention that the rank 1 case gives an explicit description of the Hilbert scheme of points on $\mathbb{C}^2$ in terms of ADHM data, and is a basic model for the computations in the higher rank cases [Baranovsky 2000; Bruzzo et al. 2011]

On $\mathbb{P}^3$, the particular rank 2 instanton bundles corresponds to the $\mathrm{SU}(2)$ gauge theoretic instantons on the four sphere S^4. Their moduli space have been studied for decades and some of its properties remained illusive for a long time. For instance, just a few years ago, its irreducibility was proved by Tikhomirov [2012; 2013], and its smoothness was shown by Jardim and Verbitsky [2014]. Recently, there has been some interest in its compactification by using torsion-free sheaves [Jardim et al. 2018; 2017b; Gargate and Jardim 2016].

In this work, we are interested in the moduli space of rank 2 framed instanton sheaves $\mathcal{M}_{\mathbb{P}^3}(c)$, on the three-dimensional projective space $\mathbb{P}^3$. More precisely we study its fixed locus $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c)$ with respect to the torus action inherited by the natural one on $\mathbb{P}^3$. We shall see that every fixed torsion-free instanton sheaf E is an extension (nontrivial in general) of ideal sheaves $\mathcal{I}_{\mathcal{C}}$ and $\mathcal{I}_{\mathcal{Z}}$, where $\mathcal{I}_{\mathcal{C}}$ is the ideal sheaf of a nonreduced Cohen–Macaulay curve $\mathcal{C}$, whose underlying reduced support is the line $l_0 = Z(z_2 = z_3 = 0)$, i.e., the unique line that is fixed by the action of $\mathbb{T}$ and does not intersect the framing line l_∞ at infinity, and $\mathcal{I}_{\mathcal{Z}}$ is the ideal sheaf of points supported on $p_0 = [1; 0; 0; 0]$ or/and $p_1 = [0; 1; 0; 0]$, in $l_0 \subset \mathbb{P}^3$. Moreover, using the fact that the double dual of such E is the trivial bundle $\mathcal{O}^2_{\mathbb{P}^3}$, we also show that every corresponding quotient $\mathcal{Q} := \mathcal{O}^2_{\mathbb{P}^3}/E$ is a pure sheaf of dimension 1 on the curve $\mathcal{C}$. These quotient sheaves $\mathcal{Q}$ are special cases of *rank* 0 *instanton sheaves* [Hauzer and Langer 2011, §6.1]. A similar phenomenon, that occurs on $\mathbb{P}^2$, is the fact that the fixed points in $\mathcal{M}_{\mathbb{P}^2}(r, c)$, under the toric action inherited from the one on $\mathbb{P}^2$, split as the sum of ideal sheaves of points, all with the same topological support given by the origin $[0; 0; 1]$ [Nakajima and Yoshioka 2005; Bruzzo et al. 2011, §3]. The difference is that the set of fixed points, in the $\mathbb{P}^3$ case, might not be isolated in general, in other words, there might be continuous families of them.

This paper is organized as follows; in Section 2, we recall the notion of *ADHM data* and their stabilities on $\mathbb{P}^3$ and how it relates to framed instanton sheaves through Horrocks monads. In Section 3 we briefly describe the inherited action, of the three-dimensional torus $\mathbb{T}$, on the ADHM data. In particular, we show that, for nonvanishing second Chern class, the fixed framed instantons are strictly torsion-free sheaves, that their double dual is trivial and that their singularity locus is pure, of dimension 1. A different proof can be found in [Gargate and Jardim 2016] for nonfixed instantons.

In Section 4, we move on to give a relation of these fixed instanton sheaves with Pandharipande–Thomas stable pairs [2009a; 2009b]. More precisely, we show that to every fixed framed rank 2 instanton sheaf $\mathcal{F}$, on $\mathbb{P}^3$, one may associate a PT-stable pair $(\mathcal{Q}, s)$. Furthermore, we show that the Euler characteristic $\chi(\mathcal{M}_{\mathbb{P}^3}(c))$ is zero, for any $c > 0$.

Section 5 is devoted to completely classify the Cohen–Macaulay supports $\mathcal{C}$ associated to the fixed PT-stable pair $(\mathcal{Q}, s)$, i.e, coming from a fixed instanton sheaf of rank 2, in $\mathbb{P}^3$. This is achieved by using results on monomial multiple structures provided by Vatne [2012].

Finally, in Section 5B, we show that a lower bound for the number of irreducible components of $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c)$, is given by the number of partitions of c. Moreover, we use results provided by Drézet [2006; 2009] in order to give an explicit description of the first canonical filtration of the rank 0 instanton sheaf $\mathcal{Q}$, for $c = 2$, and when the support is a primitive monomial double structure. When $c = 3$, we describe the 0 instanton sheaf $\mathcal{Q}$, and its filtrations for the three possible monomial structures, in particular the first nonprimitive case is given in Theorem 5.15. Using these results, we show that $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(3)$ has 7 components. For $c = 1$, we also compute the dimension of the tangent space and the obstruction at the specific fixed stable pair $(\mathcal{Q}, s)$.

We wonder whether these fixed points can arise as degenerations of locally free framed instantons, i.e., if the fixed enumerated components intersect the closure of the framed locally free instanton moduli. We think that this problem is related to reachability of sheaves, on multiple structure [Drézet 2017], and hope to address this problem in future work.

2. ADHM data and instanton sheaves

In this section we will gather useful results about ADHM data and instanton sheaves. Mostly, this material can be found in [Henni et al. 2015; Frenkel and Jardim 2008; Jardim 2006]. We consider in $\mathbb{P}^3$ the homogeneous coordinates $[z_0 : z_1 : z_2 : z_3] \in \mathbb{P}^3$ and the line ℓ_∞ given by the equations $z_0 = z_1 = 0$. Set

$$\mathrm{H}_{\mathbb{P}^1} = \langle z_0, z_1 \rangle \subset \mathrm{H}^0(\mathbb{P}^3).$$

Let V and W be complex vector spaces of dimension, respectively, c and r. Set

$$\boldsymbol{B} := \mathrm{End}(V)^{\oplus 2} \oplus \mathrm{Hom}(W, V)$$

and consider the affine spaces

$$\boldsymbol{B}_{\mathbb{P}^1} = \boldsymbol{B}_{\mathbb{P}^1}(W, V) = \boldsymbol{B}_{\mathbb{P}^1}(r, c) := \boldsymbol{B} \otimes \mathrm{H}_{\mathbb{P}^1}.$$

A point of $\boldsymbol{B}_{\mathbb{P}^1}$ will be called in this paper an *ADHM datum over* $\mathbb{P}^1$.

One can write a point of $X \in \boldsymbol{B}_{\mathbb{P}^1}$ as

$$X = (A, B, I),$$

where the above components are

$$\begin{aligned} A &= A_0 \otimes z_0 + A_1 \otimes z_1, \\ B &= B_0 \otimes z_0 + B_1 \otimes z_1, \\ I &= I_0 \otimes z_0 + I_1 \otimes z_1, \\ J &= J_0 \otimes z_0 + J_1 \otimes z_1, \end{aligned}$$

with $A_i, B_i, \in \mathrm{End}(V)$, $I_i \in \mathrm{Hom}(W, V)$ and $J_i \in \mathrm{Hom}(V, W)$, $i = 0, 1$. Hence we naturally regard $A, B \in \mathrm{Hom}(V, V \otimes \mathrm{H}_{\mathbb{P}^1})$, and also $I \in \mathrm{Hom}(W, V \otimes \mathrm{H}_{\mathbb{P}^1})$ and $J \in \mathrm{Hom}(V, W \otimes \mathrm{H}_{\mathbb{P}^1})$.

For any $P \in \mathbb{P}^1$ we define the *evaluation maps* given on generators by

$$\mathrm{ev}^1_P : \boldsymbol{B}_{\mathbb{P}^1} \to \mathbb{P}(\boldsymbol{B}), \quad X_i \otimes z_i \mapsto [z_i(P) X_i].$$

Note that $z_i(P) \in \mathbb{C}$ depends on a choice of trivialization of $\mathcal{O}_{\mathbb{P}^1}(1)$ at P but the class on projective space does not. We set $X_P := \mathrm{ev}^1_P(X)$. In particular, A_P, B_P, I_P and J_P are defined as well. For any subspace $S \subset V$, we are able to naturally well define the subspaces $A_P(S)$, $B_P(S)$, $I_P(W)$ and $\ker J_P$ of V.

We also consider the following *stability* and *costability* conditions:

Definition 2.1 [Henni et al. 2015]. Let $X = (A, B, I) \in \boldsymbol{B}_{\mathbb{P}^1}$. Let also P be a point in $\mathbb{P}^3$.

(i) X is *globally weak stable* if there is no proper subspace $S \subset V$ of dimension 1 for which hold the inclusions $A_P(S)$, $B_P(S)$, $I_P(W) \subset S$, for every $P \in \mathbb{P}^1$;

(ii) X is *globally weak costable* if there is no nonzero subspace $S \subset V$ of dimension 1 for which hold the inclusions $A_P(S), B_P(S) \subset S \subset \ker J_P$, for every $P \in \mathbb{P}^1$.

We define $\boldsymbol{B}^{\mathrm{gws}}_{\mathbb{P}^1}$ as the subsets of $\boldsymbol{B}_{\mathbb{P}^1}$ consisting of globally weak stable ADHM data over $\mathbb{P}^3$. In a similar way, we define $\boldsymbol{B}^{\mathrm{gwc}}_{\mathbb{P}^1}$. Clearly, both of them are open subsets of $\boldsymbol{B}_{\mathbb{P}^1}$, in the Zariski topology.

An *instanton sheaf* on $\mathbb{P}^n$ is a torsion free coherent sheaf E with $c_1(E) = 0$ satisfying the following cohomological conditions:

(i) For $n \geq 2$, $\mathrm{H}^0(\mathbb{P}^n, E(-1)) = \mathrm{H}^n(\mathbb{P}^n, E(-n)) = 0$.

(ii) For $n \geq 3$, $\mathrm{H}^1(\mathbb{P}^n, E(-2)) = \mathrm{H}^{n-1}(\mathbb{P}^n, E(1-n)) = 0$.

(iii) For $n \geq 4$, $\mathrm{H}^p(\mathbb{P}^n, E(-k)) = 0$ for $2 \leq p \leq n-1$, and all k.

The second Chern class $c := c_2$ is called the *charge* of E, and one can check that $c = -\chi(E) = h^1(E(-1))$. An instanton sheaf is said to be of *trivial splitting type* if there exists a line ℓ in $\mathbb{P}^3$ such that the restriction $E|_\ell$ of E on ℓ is trivial. A particular choice of trivialisation $\phi : E|_\ell \to \mathcal{O}|_\ell^{\oplus r}$ is called a *framing*, and the pair (E, ϕ) is called a *framed* instanton sheaf.

Now, we consider framed instantons on $\mathbb{P}^3$, and we fix the line ℓ_∞, given in the beginning of this section, as the framing line. Moreover, from [Henni et al. 2015, Proposition 3.6; Frenkel and Jardim 2008; Jardim 2006], we have that framed rank r instantons E of charge c are cohomologies of monads of the form

$$\mathcal{M} : V \otimes \mathcal{O}_{\mathbb{P}^3}(-1) \xrightarrow{\alpha} (V \oplus V \oplus W) \otimes \mathcal{O}_{\mathbb{P}^3} \xrightarrow{\beta} V \otimes \mathcal{O}_{\mathbb{P}^3}(1). \tag{1}$$

V is a c-dimensional vector space and can be identified with some homology group of E twisted by a some differential sheaf, via the Beilinson spectral sequence construction of the monad [Henni et al. 2015, §3]. W is r-dimensional space (this can be identified with $\mathbb{C}^r$, given a fixed basis, due to the framing). The maps α and β are given by

$$\alpha = \begin{pmatrix} A_0z_0 + A_1z_1 + \mathbb{1}z_2 \\ B_0z_0 + B_1z_1 + \mathbb{1}z_3 \\ J_0z_0 + J_1z_1 \end{pmatrix},$$

$$\beta = \big(-B_0z_0 - B_1z_1 - \mathbb{1}z_3,\ A_0z_0 + A_1z_1 + \mathbb{1}z_2,\ I_0z_0 + I_1z_1\big),$$

where $A_0, A_1, B_0, B_1 \in \mathrm{End}(V)$, $I_0, I_1 \in \mathrm{Hom}(W, V)$ and $J_0, J_1 \in \mathrm{Hom}(V, W)$. These matrices satisfy the following equations

$$\begin{gathered} [A_0, B_0] + I_0J_0 = 0, \quad [A_1, B_1] + I_1J_1 = 0, \\ [A_0, B_1] + [B_0, A_1] + I_1J_2 + I_2J_1 = 0, \end{gathered} \tag{2}$$

which are equivalent to the complex condition $\beta \circ \alpha = 0$, in the monad $\mathcal{M}$. Moreover there is a group action of $G = Gl(V)$, on the above data, given by

$$A_i \to gA_ig^{-1}, \quad B_i \to gB_ig^{-1}, \quad I_i \to gI_i, \quad J_i \to J_ig^{-1} \tag{3}$$

for $g \in G$ and $i = 0, 1$.

An easy verification, by using of additivity of the Chern character, shows that $c_3(E) = 0$.

We denote by $\mathcal{V}_{\mathbb{P}^3}(c, r)$ the space of the ADHM data satisfying (2) and in which one can define the following subvarieties, according to stabilities in Definition 2.1 $\mathcal{V}^{\mathrm{gws}}_{\mathbb{P}^3}(c, r) \subset \mathcal{V}^{\mathrm{st}}_{\mathbb{P}^3}(c, r)$. Here, we recall that subscript gws stands for globally weak stable and subscript st stands for stable, meaning that there is no subspace $S \subsetneq V$ such that $A(S), B(S), I(W) \subset S \otimes \mathrm{H}_{\mathbb{P}^1}$. The proof of the inclusion is discussed in [Henni et al. 2015, Section 2.1].

For an ADHM datum $X = (A_0, B_0, I_0, J_0, A_1, B_1, I_1, J_1)$, consider the following algebraic set

$$D_X = \{z \in \mathbb{P}^3 \mid \alpha_X \text{ is not injective}\}.$$

Note that we always have $\operatorname{codim}(D_X) \geq 2$, by the framing condition. A simpler version of [Henni et al. 2015, Propositions 3.3, 3.4] can be written as the

Theorem 2.2. *The complex* (1) *is a monad if and only if the corresponding ADHM datum is globally weak stable*, *and in this case* E, *the middle cohomology of the monad*, *is torsion-free. Moreover*, E *is a locally free framed instanton sheaf if and only if the ADHM datum* X *is globally weak costable.*

3. Torus action on the ADHM data

Now, we consider the following standard torus action of $\mathbb{T} := \mathbb{T}^3$ on $\mathbb{P}^3$ given by

$$F_t : \mathbb{T} \times \mathbb{P}^3 \to \mathbb{P}^3, \quad ((t_1, t_2, t_3), z) \mapsto [z_0; t_1 z_1; t_2 z_2; t_3 z_3].$$

This action can be lifted to the space of ADHM data as the following: let $\boldsymbol{T} := \mathbb{T} \times \widetilde{T}$, where $\widetilde{T}$ is the maximal torus of $\mathrm{GL}(W)$, given by elements of the form $e = \operatorname{diag}(e_1, \ldots, e_r)$. Let $\gamma_{e_1,\ldots,e_r}$ be the isomorphism $\mathcal{O}|_\ell^r \ni (w_1, \ldots, w_r) \to (e_1 w_1, \ldots, e_r w_r) \in \mathcal{O}|_\ell^r$. For a framed instanton sheaf $(E, \phi : E|_\ell \to \mathcal{O}|_\ell^r)$ one can define $(t, e_1, \ldots, e_r)\cdot(E, \phi) = ((F_t^{-1})^* E, \phi')$, where ϕ' is given by the composition

$$(F_t^{-1})^* E|_\ell \xrightarrow{(F_t^{-1})^*\phi} (F_t^{-1})^* \mathcal{O}|_\ell^2 \to \mathcal{O}|_\ell^r \xrightarrow{\gamma_{e_1,\ldots,e_r}} \mathcal{O}|_\ell^r.$$

Proposition 3.1. *The above action can be identified with the action on the ADHM data given by*

$$\begin{array}{llll} A_0 \to t_2 A_0, & B_0 \to t_3 B_0, & J_0 \to t_3 e J_0, & I_0 \to t_2 I_0 e^{-1}, \\ A_1 \to t_1^{-1} t_2 A_1, & B_1 \to t_1^{-1} t_3 B_1, & J_1 \to t_1^{-1} t_3 e J_1, & I_1 \to t_1^{-1} t_2 I_1 e^{-1}. \end{array} \tag{4}$$

Moreover, *the ADHM equations* (2) *and stability conditions are preserved.*

Proof. Since any framed instanton sheaf E is the middle cohomology of a monad as in (1), then the pull back $(F_t^{-1})^* E$ is the cohomology of a similar monad with maps α and β given as

$$\alpha = \begin{pmatrix} A_0 z_0 + A_1 t_1^{-1} z_1 + \mathbb{1} t_2^{-1} z_2 \\ B_0 z_0 + B_1 t_1^{-1} z_1 + \mathbb{1} t_3^{-1} z_3 \\ J_0 z_0 + J_1 t_1^{-1} z_1 \end{pmatrix},$$

$$\beta = \left(-B_0 z_0 - B_1 t_1^{-1} z_1 - \mathbb{1} t_3^{-1} z_3,\ A_0 z_0 + A_1 t_1^{-1} z_1 + \mathbb{1} t_2^{-1} z_2,\ I_0 z_0 + I_1 t_1^{-1} z_1\right).$$

Under the isomorphism

$$\begin{matrix} V\otimes\mathcal{O}_{\mathbb{P}^3} \\ \oplus \\ V\otimes\mathcal{O}_{\mathbb{P}^3} \\ \oplus \\ W\otimes\mathcal{O}_{\mathbb{P}^3}\end{matrix} \ni \begin{pmatrix} v_1 \\ v_2 \\ w\end{pmatrix} \to \begin{pmatrix} t_3^{-1}v_1 \\ t_2^{-1}v_2 \\ t_2^{-1}w\end{pmatrix}$$

the kernel of β is sent to the kernel of

$$\big(-(t_3B_0)z_0-(t_1^{-1}t_3B_1)z_1-\mathbb{1}z_3,\,(t_2A_0)z_0+(t_1^{-1}t_2A_1)z_1+\mathbb{1}z_2,\,(t_2I_0)z_0+(t_1^{-1}t_2I_1)z_1\big)$$

and the image of α is sent to the image of

$$\frac{1}{t_2t_3}\begin{pmatrix}(t_2A_0)z_0+(t_1^{-1}t_2A_1)z_1+\mathbb{1}z_2 \\ (t_3B_0)z_0+(t_1^{-1}t_3B_1)z_1+\mathbb{1}z_3 \\ (t_3J_0)z_0+(t_1^{-1}t_3J_1)z_1\end{pmatrix}$$

Composing with the action of γ_{e_1,e_2} on the framing, the assertion follows. □

Now we consider the moduli space $\mathcal{M}_{\mathbb{P}^3}(r,c):=\mathcal{V}^{st}_{\mathbb{P}^3}(r,c)/G$ (this quotient makes sense by means of [Henni et al. 2015, Section 2.3]); a datum $[X]$ is invariant under the toric action if and only if there exists an element $g_t\in G$ such that $t\cdot X=g_t\cdot X$. In other words, $[X]=[A_0,B_0,I_0,J_0,A_1,B_1,I_1,J_1]$ is $\mathbb{T}$-invariant if and only if there exists a map

$$\theta:\mathbb{T}\to G,\quad t\mapsto\theta(t)=g_t$$

such that

$$\begin{aligned} t_2A_0&=g_tA_0g_t^{-1}, & t_1^{-1}t_2A_1&=g_tA_1g_t^{-1},\\ t_3B_0&=g_tB_0g_t^{-1}, & t_1^{-1}t_3B_1&=g_tB_1g_t^{-1},\\ t_3J_0&=J_0g_t^{-1}, & t_1^{-1}t_3J_1&=J_1g_t^{-1},\\ t_2I_0&=g_tI_0, & t_1^{-1}t_2I_1&=g_tI_1.\end{aligned}\tag{5}$$

Lemma 3.2. *If $[X]$ is fixed by the torus $\mathbb{T}$, then we have $J_0=J_1=0$. Moreover, X is not globally weak costable.*

Proof. Suppose $[X]$ is fixed by the torus $\mathbb{T}$, and let $t=(t_1,t_2,t_3)$. Then one has $J_0I_0=(J_0g_t^{-1})(g_tI_0)=(t_3J_0)(t_2I_0)=t_2t_3J_0I_0$, hence $J_0I_0=0$. In the same way, one shows that $J_\alpha I_\beta=0$, for all $\alpha,\beta=0,1$. Moreover, for $A=z_0A_0+z_1A_1$ $B=z_0B_0+z_1B_1$, $I=z_0I_0+z_1I_1$ and $J=z_0J_0+z_1J_1$ such that $[A,B]+IJ=0$, for all z_0,z_1, one has

$$JBA=J[A,B]+JAB=J(-IJ)+JAB=-\underbrace{(JI)}_{0}J+JAB=JAB.$$

Thus, by induction, it follows that for any product $\widehat{C} = C_{\alpha_1} \cdot C_{\alpha_2} \cdots BA \cdots C_{\alpha_m}$, where $\alpha_i = 0, 1$, for all $i = 1, \dots, m$ and

$$C_{\alpha_i} = \begin{cases} A & \text{for } \alpha_i = 0, \\ B & \text{for } \alpha_i = 1, \end{cases}$$

one has $J\widehat{C} = C_{\alpha_1} \cdot C_{\alpha_2} \cdots BA \cdots C_{\alpha_m}$. Hence, for any such product, we have

$$J\widehat{C} = JA^l B^m, \tag{6}$$

where l and m are the numbers of A's and B's, respectively, appearing in $\widehat{C}$. On the other hand, we have

$$\begin{aligned} J_0 A_0^l B_0^m I_0 &= J_0 g_t^{-1} g_t A_0^l g_t^{-1} g_t B_0^m g_t^{-1} g_t I_0 \\ &= (t_3 J_0)(t_2^l A_0^l)(t_3^m B_0^m)(t_2 I_0) \\ &= (t_2^{m+1} t_3^{l+1}) J_0 A_0^l B_0^m I_0, \end{aligned} \tag{7}$$

for all $t \in \mathbb{T}$. Hence $J_0 A_0^l B_0^m I_0 = 0$. In the same way, it follows that

$$J_{\alpha_1} A_{\alpha_2}^l B_{\alpha_3}^l I_{\alpha_4} = 0, \tag{8}$$

for all $\alpha_i = 0, 1$. By the stability condition, we have that $V \otimes \mathrm{H}_{\mathbb{P}^1}$ is generated by the action of $C_{\alpha_1}^l C_{\alpha_2}^l$ on $I(w_1)$ and $I(w_2)$, where $\langle w_1, w_2 \rangle = \mathbb{C}^2$, Then every vector $v \in V \otimes \mathrm{H}_{\mathbb{P}^1}$ is of the form $\Sigma_{\alpha_k} C_{\alpha_1} \cdots C_{\alpha_m} I(w_1) + \Sigma_{\alpha_k} C'_{\alpha_1} \cdots C'_{\alpha_m} I(w_2)$ Hence, by (6) and (8),

$$J_\alpha v = \Sigma_{\alpha_k} J_\alpha C_{\alpha_1} \cdots C_{\alpha_m} I(w_1) + \Sigma_{\alpha_k} J_\alpha C'_{\alpha_1} \cdots C'_{\alpha_m} I(w_2) = 0.$$

Therefore, both J_0 and J_1 vanish identically. Moreover, it follows that the datum X is not globally weak costable. □

Theorem 3.3. *A $\mathbb{T}$-fixed framed rank r instanton of charge c on $\mathbb{P}^3$ is*

(i) *strictly torsion free if $c > 0$, or*

(ii) *equal to the trivial bundle if $c = 0$.*

Proof. By the correspondence in Theorem 2.2, we conclude that, for $c > 0$, the instanton E corresponding to T-fixed datum X is not locally free. From the framing, we conclude that the singularity set of the sheaf is at least 2-codimensional, hence the instanton sheaf is torsion-free, in this case. If $c = 0$, the only instanton sheaf is the trivial bundle, which is clearly fixed by the torus action. □

In the rank 2 case we have the following result:

Theorem 3.4. *Let E be a rank 2 torsion-free instanton sheaf on $\mathbb{P}^3$. Then:*

(i) *The singularity set* $\mathrm{Sing}(E)$ *of E is purely 1-dimensional.*

(ii) *Furthermore, if E is $\mathbb{T}$-fixed, then*

  (a) *its double dual E^{**} is the trivial locally free instanton sheaf $\mathcal{O}_{\mathbb{P}^3}^{\oplus 2}$, and*

  (b) *its singularity locus is topologically supported on the rational line given by $z_2 = z_3 = 0$. Moreover, the matrices A_i, B_i, for $i = 0, 1$, in the corresponding ADHM datum are nilpotent.*

Proof. Suppose E is reflexive; then E should be locally free since it is of rank two and has third Chern class $c_3(E) = 0$ [Hartshorne 1980, Proposition 2.6]. This contradicts Theorem 3.3 for $c \neq 0$. Hence the singularity set $\mathrm{Sing}(E)$ of E is 1-dimensional. it remains to check purity. This is done by showing that the quotient sheaf $\mathcal{Q} := E^{**}/E$ is pure. The sheaf $\mathcal{Q}$ is supported in codimension 2, thus we have $\mathcal{E}xt^q(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}(-4)) = 0$, for $q = 0, 1$. Moreover, by [Huybrechts and Lehn 1997, Proposition 1.1.10], $\mathcal{Q}$ is pure if, and only if, $\mathrm{codim}(\mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}(-4))) \geq 3+1 = 4$. In other words, we need to show that $\mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}(-4)))$ is the zero sheaf.

Note that $\mathcal{Q}$ is a 1-dimensional sheaf, so by [Huybrechts and Lehn 1997, Proposition 1.1.6] we have $\mathrm{codim}(\mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}(-4))) \geq 3$. Hence $\mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}(-4))$ is, supposedly, supported on a zero-dimensional subscheme in $\mathbb{P}^3$, lying inside $\mathrm{Sing}(E)$.

By Serre–Grothendieck duality can write

$$\mathrm{Ext}^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}) = \mathrm{Ext}^0(\mathcal{O}_{\mathbb{P}^3}, \mathcal{Q}(-4))^* = \mathrm{H}^0(\mathbb{P}^3, \mathcal{Q}(-4))^*.$$

Now we will show that $\mathrm{H}^0(\mathbb{P}^3, \mathcal{Q}(-4))^* = 0$. This will be achieved by using the long exact sequence in cohomology, associated to the short exact sequence

$$0 \to E(-4) \to E^{**}(-4) \to \mathcal{Q}(-4) \to 0.$$

In fact, one has

$$\mathrm{H}^0(\mathbb{P}^3, E^{**}(-4)) \to \mathrm{H}^0(\mathbb{P}^3, \mathcal{Q}(-4)) \to \mathrm{H}^1(\mathbb{P}^3, E(-4)). \tag{9}$$

Claim 1. $\mathrm{H}^1(\mathbb{P}^3, E(-4)) = 0$.

From the monad associated to E, one has the sequences

$$0 \to \ker\beta \to \mathcal{O}_{\mathbb{P}^3}^{\oplus(2c+2)} \xrightarrow{\beta} \mathcal{O}_{\mathbb{P}^3}(1)^{\oplus c} \to 0, \tag{10}$$

$$0 \to \mathcal{O}_{\mathbb{P}^3}(-1)^{\oplus c} \to \ker\beta \to E \to 0. \tag{11}$$

By using (10) and its dual, one can verify that $\mathrm{H}^1(\mathbb{P}^3, \ker\beta \otimes \mathcal{O}_{\mathbb{P}^3}(-4)) = 0$ and $\mathrm{H}^0(\mathbb{P}^3, (\ker\beta)^*(-4)) = 0$.

Twisting the sequence (11) by $\mathcal{O}_{\mathbb{P}^3}(-4)$, and analysing the long sequence in cohomology, it is not difficult to check that $\mathrm{H}^1(\mathbb{P}^3, E(-4)) = 0$.

Claim 2. $\mathrm{H}^0(\mathbb{P}^3, E^{**}(-4)) = 0$.

On the other hand, by dualizing this sequence, one gets

$$0 \to E^{**}(-4) \to (\ker\beta)^* \otimes \mathcal{O}_{\mathbb{P}^3}(-4) \to \mathcal{O}_{\mathbb{P}^3}(-3)^{\oplus c} \to \mathcal{E}xt^1(E, \mathcal{O}_{\mathbb{P}^3}(-4)) \to 0,$$

where we used the fact that $E^{**} \cong E^*$, since E^* is a rank 2 reflexive sheaf on $\mathbb{P}^3$, with trivial first Chern class. Breaking this sequence into

$$0 \to E^{**}(-4) \to (\ker\beta)^* \otimes \mathcal{O}_{\mathbb{P}^3}(-4) \to J \to 0$$

and using the long exact sequence in cohomology one gets the desired result, since $\mathrm{H}^0(\mathbb{P}^3, (\ker\beta)^*(-4))$ is trivial.

It follows from the proved claims and the sequence (9) that $\mathrm{Ext}^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}) = 0$, as desired. On the other hand, from the local-to-global spectral sequence one has

$$\mathrm{Ext}^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3}) = \mathrm{H}^0(\mathbb{P}^3, \mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3})) \oplus \mathrm{H}^1(\mathbb{P}^3, \mathcal{E}xt^2(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3})).$$

Both contributing terms

$$\mathrm{H}^0(\mathbb{P}^3, \mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3})), \quad \mathrm{H}^1(\mathbb{P}^3, \mathcal{E}xt^2(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3})),$$

to the local-to-global spectral sequence must be trivial. Finally, observe that $\dim \mathrm{H}^0(\mathbb{P}^3, \mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3})) = 0$ is the length of the sheaf $\mathcal{E}xt^3(\mathcal{Q}, \mathcal{O}_{\mathbb{P}^3})$, which must be zero since any sheaf supported on a zero-dimensional subscheme of $\mathbb{P}^3$, with zero length is the zero sheaf. Hence $\mathcal{Q}$ is pure.

To see that $c_3(E^{**}) = 0$, it suffices to show the following:

Claim 3. *E^{**} is the cohomology of a monad.*

Clearly, E^{**} satisfies $\mathrm{H}^0(\mathbb{P}^3, E(-1)) = \mathrm{H}^3(\mathbb{P}^3, E(-3)) = 0$, since it is also framed. Moreover, it also satisfies $\mathrm{H}^2(\mathbb{P}^3, E^{**}(-2)) = 0$, since this sits in the sequence

$$\mathrm{H}^2(\mathbb{P}^3, E(-2)) \to \mathrm{H}^2(\mathbb{P}^3, E^{**}(-2)) \to \mathrm{H}^2(\mathbb{P}^3, \mathcal{Q}(-2)),$$

with $\mathrm{H}^2(\mathbb{P}^3, E(-2)) = 0$, by instanton definition (page 44), and $\mathrm{H}^2(\mathbb{P}^3, \mathcal{Q}(-2)) = 0$ from the fact that $\dim \mathrm{Supp}(\mathcal{Q}) = 1$.

The dual of the complex (1) is a perverse instanton sheaf of trivial splitting type [Henni et al. 2015, Section 5.4; Hauzer and Langer 2011] whose zeroth cohomology is E^*, and first cohomology is $\mathcal{E}xt^1(E, \mathcal{O}_{\mathbb{P}^3})$. Then by [Henni et al. 2015, Theorem 5.13], the hypercohomology $\mathbb{H}^1(\mathbb{P}^3, \mathcal{M}^*(-2))$ vanishes, and since this is just $\mathrm{H}^1(\mathbb{P}^3, E(-2)) \oplus \mathrm{H}^0(\mathbb{P}^3, \mathcal{E}xt^1(E^*, \mathcal{O}_{\mathbb{P}^3})(-2))$, one gets in particular that $\mathrm{H}^1(\mathbb{P}^3, E^*(-2)) = 0$. Again, one has $E^{**} \cong E^*$; hence the double dual satisfies the definition of instanton sheaf, and this proves the claim.

Finally, recall that the double dual of any sheaf is reflexive. Thus E^{**} is a reflexive framed instanton sheaf, which fixed by the torus action on $\mathbb{P}^3$. Moreover, by [Hartshorne 1980, Proposition 2.6], E^{**} should be locally free, since $c_3(E^{**}) = 0$. But according to Theorem 3.3 we should have $E^{**} = \mathcal{O}_{\mathbb{P}^3}^{\oplus 2}$, since the nontrivial $\mathbb{T}$-fixed instantons should be strictly torsion-free.

To conclude the proof of item (b) we note that E is torsion free, by Theorem 3.3, and its singularity set is purely 1-dimensional, by item (i). Moreover, by the framing condition it does not intersect the framing line. In particular, the singularity set is also $\mathbb{T}$-invariant. But the only invariant codimension 2 subscheme of $\mathbb{P}^3$, as a toric variety, which does not intersect the framing line is supported on the line $[z_0, z_1, 0, 0]$.

The singularity set is the locus on which the map α, in the monad (1), is not injective. In particular, one can characterize it in terms of the eigenvalues equations

$$\det[(A_0 z_0 + A_1 z_1) + z_2 \mathbb{1}] = 0, \quad \det[(B_0 z_0 + B_1 z_1) + z_3 \mathbb{1}] = 0.$$

But we just showed that all the corresponding eigenvalues z_2, z_3 must be 0. Hence the matrices $(A_0 z_0 + A_1 z_1)$ and $(B_0 z_0 + B_1 z_1)$ must be nilpotent, for all z_0, z_1, and consequently, the result follows. □

We note that a proof for item (i) can also be found in [Gargate and Jardim 2016], however, we gave our own proof for completeness. Moreover, this is shorter version concerned mainly with $\mathbb{T}$-fixed locus.

From the above, we see that if $[X] \in \mathcal{M}_{\mathbb{P}^3(r,c)}$ is a $\mathbb{T}$-fixed point, then it is represented by a datum $X = (A_0, B_0, I_0, A_1, B_1, I_1)$ satisfying the equations

$$[A_0, B_0] = 0, \quad [A_1, B_1] = 0, \quad [A_0, B_1] + [B_0, A_1] = 0. \tag{12}$$

4. Quotients and PT-stable pairs

In this section we will adopt the following viewpoint: let E be a $\mathbb{T}$-invariant torsion-free instanton sheaf of rank 2, then E fits in the short exact sequence

$$0 \to E \to \mathcal{O}_{\mathbb{P}^3}^2 \to \mathcal{Q} \to 0.$$

The Hilbert polynomials of sheaves involved in this sequence are

$$P_E(m) = \tfrac{1}{3} m^3 + 2m^2 + \left(\tfrac{11}{3} - c\right) m + (2 - 2c), \quad P_{\mathcal{Q}}(m) = cm + 2c.$$

Since every such sheaf E is given by a datum $X \in \mathcal{V}_{\mathbb{P}^3}^{\mathbb{T}}(c)$, one can think of $\mathcal{M}_{\mathbb{P}^3}^{\mathbb{T}}(c)$ as an open subset of the scheme

$$\boldsymbol{Quot}^{[cm+2c]}_{\mathcal{O}_{\mathbb{P}^3}^2, [l_0]},$$

which parametrises quotients $\mathcal{O}_{\mathbb{P}^3}^2 \twoheadrightarrow \mathcal{Q}$ with $\mathcal{Q}$ is $\mathbb{T}$-fixed 1-dimensional pure sheaf, topologically supported on the fixed line $l_0 : \mathbb{P}^1 \hookrightarrow \mathbb{P}^3$, not intersecting the framing line l_∞, that is l_0 is given by $[z_0; z_1] \mapsto [z_0; z_1; 0; 0]$. The instanton cohomological conditions on E imply the $\mathcal{Q}$ should satisfy $\mathrm{H}^0(\mathbb{P}^3, \mathcal{Q}(-2)) = \mathrm{H}^1(\mathbb{P}^3, \mathcal{Q}(-2)) = 0$. Obviously these are open conditions in flat families by semicontinuity.

Recall that a *rank* 0 *instanton sheaf* is a pure sheaf of codimension 2 satisfying $\mathrm{H}^0(\mathbb{P}^3, \mathcal{Q}(-2)) = \mathrm{H}^1(\mathbb{P}^3, \mathcal{Q}(-2)) = 0$. This definition was introduced in [Hauzer and Langer 2011, §6.1].

Lemma 4.1 [Jardim et al. 2017a, Lemma 6]. *Every rank* 0 *instanton sheaf is μ-semistable.*

Proof. Let $\boldsymbol{T}$ a subsheaf of $\mathcal{Q}$ with Hilbert polynomial $P_{\boldsymbol{T}}(m) = am + b$. Note that $\mathrm{H}^0(\mathbb{P}^3, \mathcal{Q}(-2)) = 0$ implies that $\mathrm{H}^0(\mathbb{P}^3, \boldsymbol{T}(-2)) = 0$. Thus $P_{\boldsymbol{T}}(-2) = -2a + b = -\mathrm{H}^1(\mathbb{P}^3, \boldsymbol{T}(-2)) \leq 0$. Hence $\mu(\boldsymbol{T}) = \frac{b}{a} \leq 2 = \frac{2c}{c} = \mu(\mathcal{Q})$. □

Thus the quotient $\mathcal{Q}$ is a rank 0 instanton sheaf, and hence μ-semistable.

Following [Pandharipande and Thomas 2009b], let $q \in \mathbb{Q}[x]$ a degree 1 polynomial with positive leading coefficient. For $n \in \mathbb{Z}$ and $\beta \in \mathrm{H}^2(\mathbb{P}^3, \mathbb{Z})$, we will consider pairs $\mathcal{O}_{\mathbb{P}^3} \xrightarrow{s} \mathcal{Q}$, on $\mathbb{P}^3$, where $\mathcal{Q}$ is a pure sheaf, of dimension 1 on $\mathbb{P}^3$, with Hilbert polynomial

$$\chi(\mathcal{Q}(m)) = m \cdot \beta + n.$$

The polynomial q is viewed as a stability parameter, and s is a nonzero section. We also let $r_{\mathcal{T}}$ denote, for any sheaf $\mathcal{T}$, the leading coefficient of its Hilbert polynomial. Since $\mathcal{Q}$ is pure, then any proper subsheaf $\mathcal{T}$ of $\mathcal{Q}$ is also pure of the same dimension. Therefore $r_{\mathcal{T}} > 0$. We say that the pair $(\mathcal{Q}, s)$ is *q-semistable* if, for any proper subsheaf $\mathcal{T} \subset \mathcal{Q}$, the inequality

$$\frac{\chi(\mathcal{T}(m))}{r_{\mathcal{T}}} \leq \frac{\chi(\mathcal{Q}(m)) + q(m)}{r_{\mathcal{Q}}}, \qquad m \gg 0 \tag{13}$$

holds, and for any proper subsheaf $\mathcal{T} \subset \mathcal{Q}$, through which the section s factors, the inequality

$$\frac{\chi(\mathcal{T}(m)) + q(m)}{r_{\mathcal{T}}} \leq \frac{\chi(\mathcal{Q}(m)) + q(m)}{r_{\mathcal{Q}}}, \qquad m \gg 0 \tag{14}$$

holds. We say the pair is *q-stable* if these inequalities are strict. The moduli space of such q-(semi)stable pairs is denoted by $P_n^q(\mathbb{P}^3, \beta)$, and was constructed by Le Potier [1995], using GIT. Moreover the pair $(\mathcal{Q}, s)$ is said to be *stable* if it is stable in the large q limit, i.e., for sufficiently large coefficients of q. In this case, we will drop the superscript q and denote the moduli of such stable pairs, simply, by $P_n(\mathbb{P}^3, \beta)$. For more details about its construction and the fact that it has a perfect obstruction, and hence a well defined virtual class, one might consult [Pandharipande and Thomas 2009b].

We also recall the following:

Lemma 4.2 [Pandharipande and Thomas 2009b, Lemma 1.3]. *A pair* $(\mathcal{Q}, s)$ *is stable* (*in the large* q *limit*) *if*, *and only if*,

(i) $\mathcal{Q}$ *has pure dimension* 1,

(ii) *the cokernel of* s *has dimension zero.*

In what follows, we will say that $(\mathcal{Q}, s)$ is a *stable* 0 *instanton pair* if it is stable (in the large q limit) and $\mathcal{Q}$ is a rank 0 instanton sheaf. Moreover, as in [Jardim et al. 2017a, p. 402], an instanton sheaf E will be called *quasitrivial* if its double dual is the trivial sheaf.

Recall also that if $\mathcal{Q}$ is associated to a quasitrivial torsion-free instanton, then one has the following commutative diagram:

$$
\begin{array}{ccccccccc}
 & & & & 0 & & & & \\
 & & & & \downarrow & & & & \\
 & & & & \mathcal{O}_{\mathbb{P}^3} & \overset{s}{\searrow} & & & \\
 & & & & \downarrow & & & & \\
0 & \longrightarrow & E & \longrightarrow & \mathcal{O}^2_{\mathbb{P}^3} & \longrightarrow & \mathcal{Q} & \longrightarrow & 0 \\
 & & & & \downarrow & & & & \\
 & & & & \mathcal{O}_{\mathbb{P}^3} & & & & \\
 & & & & \downarrow & & & & \\
 & & & & 0 & & & &
\end{array}
\tag{15}
$$

This define a section s of $\mathcal{Q}$. Let us put $\mathcal{G} := \operatorname{Im}(s)$ and $\mathcal{I} := \ker(s)$. Then $\mathcal{I}$ is an ideal sheaf in $\mathcal{O}_{\mathbb{P}^3}$ of a subscheme S of pure dimension 1 in $\mathbb{P}^3$, with structure sheaf $\mathcal{O}_{\mathcal{C}} = \mathcal{G}$. Moreover, if E is $\mathbb{T}$-fixed, then $\mathcal{Q}$ is a $\mathbb{T}$-fixed rank 0 instanton sheaf. It follows from Theorem 3.4 that the theoretical support of S is exactly the line l_0, defined by the locus $(z_2 = z_3 = 0)$, in $\mathbb{P}^3$.

Proposition 4.3. *Let* $(\mathcal{Q}, s)$ *be* 0 *instanton pair associated to a rank* 2 *quasitrivial instanton sheaf* E *on* $\mathbb{P}^3$. *Then*:

(i) $(\mathcal{Q}, s)$ *is stable.*

(ii) *Moreover*, *if* E *is framed and* $\mathbb{T}$*-fixed*, *then* $E \in \operatorname{Ext}^1(\mathcal{I}_{\mathcal{Z}}, \mathcal{I}_{\mathcal{C}})$, *where* $\mathcal{I}_{\mathcal{C}}$ *is a* $\mathbb{T}$*-fixed ideal sheaf of a subscheme of pure dimension* 1 *and* $\mathcal{I}_{\mathcal{Z}}$ *is* $\mathbb{T}$*-fixed ideal sheaf of a zero-dimensional subscheme and neither* $\mathcal{Z}$, *nor* $\mathcal{C}$, *intersects the line* l_∞.

Proof. (i) By [Jardim et al. 2017a, Corollary 5], it follows that $\operatorname{coker} s$ is zero-dimensional and since $\mathcal{Q}$ is pure, the result follows by Lemma 4.2.

(ii) First, notice that since $\mathcal{Q}$ is a rank 0 instanton, one can complete the commutative diagram (15) to get

$$\begin{array}{ccccccccc} & & 0 & & 0 & & 0 & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathcal{I}_{\mathcal{C}} & \longrightarrow & \mathcal{O}_{\mathbb{P}^3} & \longrightarrow & \mathcal{O}_{\mathcal{C}} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E & \longrightarrow & \mathcal{O}_{\mathbb{P}^3}^2 & \longrightarrow & \mathcal{Q} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & \mathcal{I}_{\mathcal{Z}} & \longrightarrow & \mathcal{O}_{\mathbb{P}^3} & \longrightarrow & \mathcal{Z} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & & 0 & & 0 & & \end{array} \tag{16}$$

where $\mathcal{O}_{\mathcal{C}} := \mathrm{Im}(s)$, $\mathcal{I}_{\mathcal{C}} := \ker(s)$ and $\mathcal{Z} := \mathrm{coker}(s)$. By [Jardim et al. 2017a, Corollary 5] it follows that the rank 2 instanton sheaf E belongs to $\mathrm{Ext}^1(\mathcal{I}_{\mathcal{Z}}, \mathcal{I}_{\mathcal{C}})$, where $\mathcal{I}_{\mathcal{C}}$ is an ideal sheaf of a subscheme in of pure dimension 1 and $\mathcal{I}_{\mathcal{Z}}$ is ideal sheaf of a zero-dimensional subscheme in $\mathbb{P}^3$. Furthermore, if E is $\mathbb{T}$-fixed, then so are $\mathcal{I}_{\mathcal{Z}}$ and $\mathcal{I}_{\mathcal{C}}$, since $\mathcal{Z}$ and $\mathcal{C}$ are fixed. Finally, $\mathcal{C}$ and $\mathrm{Supp}(\mathcal{Z})$ are supported on $\mathrm{Sing}(E)$, and by item (b) of Theorem 3.4, $\mathrm{Sing}(E)$ has vacuous intersection with the framing line l_∞, so the result follows. □

Theorem 4.4. *The Euler characteristic of $\mathcal{M}_{\mathbb{P}^3}(c)$ is $\chi(\mathcal{M}_{\mathbb{P}^3}(c)) = 0$, for all $c > 0$. Moreover if $c = 1$, then the Poincaré Polynomial of $\mathcal{M}_{\mathbb{P}^3}(1)$ is*

$$\mathcal{P}_{\mathcal{M}_{\mathbb{P}^3}(1)}(t) = \sum_{i=0}^{13}(1 - \delta_{1,i} - \delta_{12,i})t^i$$

Before giving a proof, we recall some useful definitions mostly from [Serre 1958]:

Definition 4.5. Let Y be an algebraic space endowed with a right (or left) action of a group G, and let $\pi : Y \to X$ be a morphism from Y to the algebraic space X. We call the triple (G, Y, X) (or just Y) a fibered system if π satisfies $\pi(x \cdot g) = \pi(x)$ for all $x \in X$ and $g \in G$.

A fibered system Y is said to be *locally trivial* if for every Zariski open $U \subset X$, the restriction $Y|_U$, of Y on U, is isomorphic to $U \times G$, with the endowed operations $(x, g)g' = (x, gg')$ and the canonical projection $U \times G \to U$, and is said to be *locally isotrivial* if for every open $U \subset X$ there is an unramified morphism $f : U' \to U$ over U such that the inverse image $f^{-1}Y|_U$, of $Y|_U$, is trivial. Y is said to be *trivial* if Y is isomorphic to $X \times G$.

A group G is said to be *special* if every locally isotrivial fibered system (G, Y, X) is locally trivial. Finally, an isotrivial fibered system (G, Y, X) is called a *G-principal fibration.* If, moreover, the morphism π is flat and (G, Y, X) is locally isotrivial, then (G, Y, X) (or just Y) is called a *G-principal bundle*.

Proof of Theorem 4.4. Let $\mathcal{I}(c)$ is the moduli space of rank 2 locally free instantons on $\mathbb{P}^3$ (without framing). We first remark that, for all $c > 0$, $\mathcal{M}_{\mathbb{P}^3}(c)$ is an $\mathrm{Sl}(2, \mathbb{C})$-bundle over $\bar{\mathcal{I}}(c)$, where the projection is given by forgetting the framing. Since the group $\mathrm{Sl}(2, \mathbb{C})$ is special [Grothendieck 1958], in the sense above, we have that every G-principal bundle is locally trivial in the Zariski topology. In particular $\mathcal{M}_{\mathbb{P}^3}(c) \to \bar{\mathcal{I}}(c)$ is a locally trivial $\mathrm{Sl}(2, \mathbb{C})$-principal bundle. Hence, one can write the Poincaré polynomial[1] of $\mathcal{M}_{\mathbb{P}^3}(c)$ as

$$\mathcal{P}_{\mathcal{M}_{\mathbb{P}^3}(c)}(t) = \mathcal{P}_{\bar{\mathcal{I}}(c)}(t) \times \mathcal{P}_{\mathrm{Sl}(2,\mathbb{C})}(t),$$

and since $\mathrm{Sl}(2, \mathbb{C}) \approxeq_{\mathrm{diff}} \mathrm{SU}(2) \times \mathbb{R}^3$, one gets $\mathcal{P}_{\mathrm{Sl}(2,\mathbb{C})}(t) = 1 + t^3$. By putting $t = -1$, it follows that $\chi(\mathcal{M}_{\mathbb{P}^3}(c)) = 0$.

In [Jardim et al. 2017b, Section 6], the authors prove that $\bar{\mathcal{I}}(1) \cong \mathbb{P}^5$. Hence, for $c = 1$ the Poincaré polynomial is computed from the product formula. □

5. Relation with multiple structures

In this section we explore the relation of the rank 0 instanton sheaves and sheaves on multiple structures [Vatne 2012; Drézet 2006; 2009; Nollet 1997]. This allows us to give a concrete description in the lower charge cases $c = 1, 2$, as well as the multiple primitive cases (see Section 5B). Moreover, we use such a description to compute the Euler characteristic of $\mathcal{M}_{\mathbb{P}^3}(1)$. We also give a lower bound on the number of irreducible components.

5A. ***Monomial multiple structures.*** Most of the material in this subsection is borrowed from [Vatne 2012], with the assumption that the ambient space is $\mathbb{P}^3$. Let $i : X = \mathbb{P}^1 \to \mathbb{P}^3$ be a linear subspace with saturated ideal $\mathcal{I}_X$, $X^{(i)} \subset \mathbb{P}^3$ the i-th infinitesimal neighborhood of X, with ideal $(\mathcal{I}_X)^{i+1}$, and Y a Cohen–Macaulay multiple structure with $Y_{\mathrm{red}} = X$, whose ideal is generated by monomials. Then the following filtration of Y exists:

$$X = Y_0 \subset Y_1 \subset \cdots \subset Y_{k-1} \subset Y_k = Y; \quad Y_i = Y \cap X^{(i)}, \tag{17}$$

for some k, and every term Y_i is also Cohen–Macaulay since X is a Cohen–Macaulay curve [Vatne 2002, Corollary 2.6]. If $\mathcal{I}_i$ is the ideal sheaf of Y_i, then there are two short exact sequences

$$0 \to \mathcal{I}_{i+1}/\mathcal{I}_X\mathcal{I}_i \to \mathcal{I}_i/\mathcal{I}_X\mathcal{I}_i \to \mathcal{L}_i \to 0$$

and

$$0 \to i_*\mathcal{L}_i \to \mathcal{O}_{Y_{i+1}} \to \mathcal{O}_{Y_i} \to 0.$$

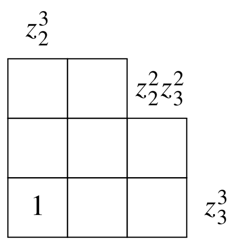
[1]For the multiplicative property of the Poincaré polynomial the reader might see [Brion and Peyre 2002, Introduction], for instance.

The first exact sequence define the locally free $\mathcal{O}_X$-modules $\mathcal{L}_j$; see [Nollet 1997] for more details.

One important result that will be used is the following:[2]

Proposition 5.1 [Vatne 2012, Proposition 1]. *There is a bijective (inclusion reversing) correspondence between Cohen–Macaulay monomial ideals in two variables and Young diagrams. Under this bijection, the number of boxes in the Young diagram is the multiplicity of the scheme defined by the corresponding ideal and whose reduced structure is a fixed line in* $\mathbb{P}^3$.

For instance, if we choose $\mathcal{I}_X = \langle z_2, z_3 \rangle \subset S := \mathbb{C}[z_0, z_1, z_2, z_3]$, the Cohen–Macaulay monomial ideal $J := \langle z_2^3, z_2^2 z_3^2, z_3^3 \rangle$ will correspond to the diagram

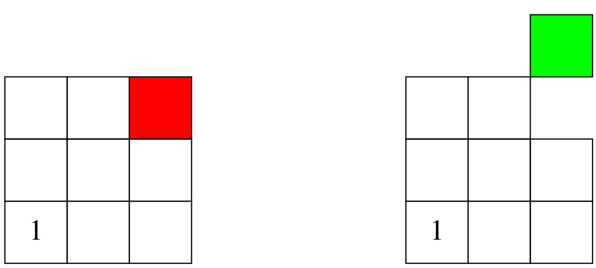


The number of boxes being 8, we have that J is an ideal of a Cohen–Macaulay multiplicity 8 structure on the line X. The line X itself corresponds to the box $\boxed{1}$.

Definition 5.2.

- An *inner box* of a Young diagram will mean a box not in the diagram but such that the box below it and the box to its left are both in the diagram.
- An *outer box* of a Young diagram will mean a box not in the diagram and such that the box below it and the box to its left are both outside the diagram, but its lower left angle touches a box in the diagram.

Example 5.3. In the diagram associated to $J := \langle z_2^3, z_2^2 z_3^2, z_3^3 \rangle$, above, the red box is *inner*, while the green box is *outer*:



[2] We only need, for our purpose, this restricted version of the more general result proved by Vatne.

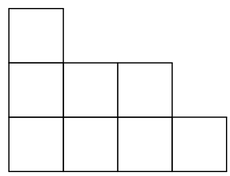
Proposition 5.4 [Vatne 2012, Proposition 4]. *Given a Cohen–Macaulay monomial ideal I with support on a line in $\mathbb{P}^3$, and its corresponding Young diagram T. Then I fits in the exact sequence*

$$0 \to \bigoplus_j \mathcal{O}_{\mathbb{P}^3}(-n_{2i}) \to \bigoplus_i \mathcal{O}_{\mathbb{P}^3}(-n_{1i}) \to I \to 0,$$

where n_{1i} is the weight of the i-th inner box and n_{2j} is the weight of the j-th inner box, for some chosen indices i of inner boxes and j of outer boxes in T.

This way, the syzygies correspond to the outer boxes.

Example 5.5. For the ideal I corresponding to the Young diagram



one has four inner boxes with weights $n_{11} = 3$, $n_{12} = 3$, $n_{13} = 4$, $n_{14} = 4$, and three outer boxes with weights $n_{21} = 4$, $n_{22} = 5$, $n_{23} = 5$. Hence, one gets

$$0 \to \mathcal{O}_{\mathbb{P}^3}(-4) \oplus \mathcal{O}_{\mathbb{P}^3}(-5)^{\oplus 2} \to \mathcal{O}_{\mathbb{P}^3}(-4)^{\oplus 2} \oplus \mathcal{O}_{\mathbb{P}^3}(-3)^{\oplus 2} \to I \to 0,$$

Theorem 5.6. *For a $\mathbb{T}$-fixed stable 0 instanton pair $(\mathcal{Q}, s)$ of charge c:*

(i) *The associated scheme $\mathcal{C}$ is a multiple structure that corresponds to a Young diagram T of weight c. Moreover if the Young diagram is of the form*[3] $\nu = (\nu_1 \geq \nu_2 \geq \cdots \geq \nu_k)$ *of c, then $\mathcal{C}$ has Hilbert polynomial*

$$\chi_{\mathcal{C}}(m) := \chi(\mathcal{O}_{\mathcal{C}}(m)) = cm + 3c - \sum_{i=1}^{k} \frac{\nu_i(\nu_i + 2i + 1)}{2},$$

and $\mathcal{I}_{\mathcal{C}}$ is a smooth point in its Hilbert scheme of closed subschemes of $\mathbb{P}^3$. The dimension of the Hilbert scheme, of subschemes of $\mathbb{P}^3$, at $\mathcal{I}_{\mathcal{C}}$ is given by

$$D_{\mathcal{I}_{\mathcal{C}}} = \sum_{n_{2j} \geq n_{1i}} \binom{n_{2j} - n_{1i} + 3}{3} + \sum_{n_{1i} \geq n_{2j}} \binom{n_{1i} - n_{2j} + 3}{3} - \sum_{n_{2j} \geq n_{2i}} \binom{n_{2j} - n_{2i} + 3}{3} - \sum_{n_{1j} \geq n_{1i}} \binom{n_{1j} - n_{1i} + 3}{3} + 1.$$

[3]The Young diagram is associated to a partition ν of c, where the i-th column represents the i-th part ν_i, $i = 1, \ldots, k$.

(ii) *Moreover, the associated sheaf* $\mathcal{Z} := \operatorname{coker}(s)$ *has length*

$$l_{\mathcal{Z}} = \frac{1}{2}\sum_{i=1}^{k} \nu_i^2 + \sum_{i=1}^{k} i\nu_i - \frac{c}{2},$$

where $\nu = (\nu_1 \geq \nu_2 \geq \cdots \geq \nu_k)$ *is the partition of* c, *represented by the Young diagram* T *associated to the multiple structure* $\mathcal{C}$.

Proof. (i) For a $\mathbb{T}$-fixed stable 0 instanton pair $(\mathcal{Q}, s)$ the schematic support $\mathcal{C}$ of $\mathcal{Q}$ should also be invariant. By Theorem 3.4, it follows that it is a multiple structure on the unique line that does not intersect the framing line in $\mathbb{P}^3$. Hence its ideal $\mathcal{I}_{\mathcal{C}}$ should be generated by monomials. The Hilbert polynomial $\chi_{\mathcal{C}}(m)$ can be computed according to [Vatne 2012, Corollary 2] and using the fact that the weight of a box (i, j) is given by $w_{i,j} = i + j - 2$. The dimension $D_{\mathcal{I}_{\mathcal{C}}}$, of the Hilbert scheme of subschemes of $\mathbb{P}^3$ at $\mathcal{C}$, follows from [Vatne 2012, Corollary 1].

(ii) The statement about $\mathcal{Z}$ follows from item (i) and the exact sequence

$$0 \to \mathcal{O}_{\mathcal{C}} \to \mathcal{Q} \to \mathcal{Z} \to 0. \qquad \square$$

The above result classifies all scheme theoretic supports of the $\mathbb{T}$-fixed stable 0 instanton pair $(\mathcal{Q}, s)$.

5B. ***Sheaves on multiple structures.*** After classifying the possible schematic supports of the pair, we will now study the sheaf $\mathcal{Q}$, emanating from the $\mathbb{T}$-fixed stable pair $(\mathcal{Q}, s)$, as a sheaf on the monomial double structure $\mathcal{C}$ defined over the line $l_0 = (z_2 = z_3 = 0)$. For instance, we give a complete description of $\mathcal{Q}$ for small charges, namely $c = 1, 2$ and 3. In order to achieve this goal we first recall some results from [Drézet 2006; 2009].

For $X = l_0 \subset Y \subset \mathbb{P}^3$, as in Section 5A, with a filtration (17) we say that Y is *primitive* if for every $x \in X$, there exists a surface S of $\mathbb{P}^3$ which is smooth at x and containing a neighborhood of x in Y. In this case, $L = \mathcal{I}_X/\mathcal{I}_{Y_2}$ is an invertible sheaf on X and we have $\mathcal{I}_{Y_i}/\mathcal{I}_{Y_{i+1}} = L^i$ for $1 \leq i \leq c$. This means that for a point $x \in X$, there are elements z_2, z_3, t, of the maximal ideal $m_{X,x}$ of x in $\mathcal{O}_{X,x}$, such that their images in $m_{X,x}/m_{X,x}^2$ form a basis and for all $1 \leq i \leq c$ one has $\mathcal{I}_{Y_i,x} = \langle z_2, z_3^i \rangle$. Let $\mathcal{F}$ be a coherent sheaf over Y.

Definition 5.7. The *first canonical filtration* of $\mathcal{F}$ is the filtration

$$\mathcal{F}_{c+1} = 0 \subset \mathcal{F}_c \subset \cdots \subset \mathcal{F}_2 \subset \mathcal{F}_1 = \mathcal{F}$$

such that, for $1 \leq i \leq c$, $\mathcal{F}_{i+1}$ is inductively defined as the kernel of the restriction morphism $\mathcal{F}_i \to \mathcal{F}_i|_X$.

In this way one has $\mathcal{F}_i/\mathcal{F}_{i+1} = \mathcal{F}_i|_X$ and $\mathcal{F}/\mathcal{F}_{i+1} = \mathcal{F}|_{Y_i}$. The graded object $\mathrm{Gr}(\mathcal{F}) = \bigoplus_{i=1}^{c} \mathcal{F}_i/\mathcal{F}_{i+1}$ is then an $\mathcal{O}_X$-module.[4]

Some properties of these filtrations can be listed as follows [Drézet 2009, §3]:

- For the ideal $\mathcal{I}_X$, of X, in $\mathcal{O}_Y$ and a coherent sheaf $\mathcal{F}$, over Y, one has $\mathcal{F}_i = \mathcal{I}_X^i\mathcal{F}$ so that $\mathrm{Gr}(\mathcal{F}) = \bigoplus_{i=0}^{c-1} \mathcal{I}_X^i\mathcal{F}/\mathcal{I}_X^{i+1}\mathcal{F}$.
- $\mathcal{F}_i = 0$ if and only if $\mathcal{F}$ is a sheaf over Y_i.
- For each $0 \leq i \leq c$, $\mathcal{F}_i$ is a coherent sheaf over Y_i with first canonical filtration $0 \subset \mathcal{F}_c \subset \cdots \subset \mathcal{F}_{i+1} \subset \mathcal{F}_i$.
- Morphisms of coherent sheaves $\mathcal{F} \to \mathcal{G}$, on Y, induce morphisms of first canonical filtrations $\mathcal{F}_i \to \mathcal{G}_i$, for all $0 \leq i \leq c$, and hence induce morphisms of the graded objects $\mathrm{Gr}(\mathcal{F}) \to \mathrm{Gr}(\mathcal{G})$.

Definition 5.8.

- *The generalised rank* is defined by the integer $R(\mathcal{F}) = \mathrm{rk}(\mathrm{Gr}(\mathcal{F}))$.
- *The generalised degree* is defined by the integer $\mathrm{Deg}(\mathcal{F}) = \deg(\mathrm{Gr}(\mathcal{F}))$.

The generalised rank and degree are defined so that they behave additively on exact sequence on Y. In general the usual rank and degree fail to satisfy this condition. Moreover we have the following generalised Riemann–Roch Theorem:

Theorem 5.9 [Drézet 2006, Theorem 4.2.1]. *For a coherent sheaf $\mathcal{F}$, over Y, we have*

$$\chi(\mathcal{F}) = \mathrm{Deg}(\mathcal{F}) + R(\mathcal{F})(1 - g_Y).$$

Here, g_Y is the genus of the curve Y.

5B1. *Stable rank* 0 *instanton pair of charge* 1. In this case the only possible support is the line l_0, the line that does not intersect the framing line l_∞. The sheaf $\mathcal{Q}$ sits in the short exact sequence

$$0 \to \mathcal{O}_{l_0} \to \mathcal{Q} \to \mathcal{Z} \to 0,$$

where $\mathcal{Z}$ is the structure sheaf of one point. Hence the only possibilities are $\mathcal{Q} = \mathcal{O}_{l_0}(p_i)$, $i = 0, 1$ Where $p_0 = [1; 0; 0; 0]$ and $p_1 = [0; 1; 0; 0]$. We point out that the rank 2 fixed instanton bundles given by $\ker(\mathcal{O}_{\mathbb{P}^3}^2 \to \mathcal{O}_{l_0}(p_i))$ are null correlation sheaves [Ein 1982]. Moreover, we see that these $\mathbb{T}$-fixed points are isolated.

Corollary 5.10. *The moduli $\mathcal{M}_{\mathbb{P}^3}(1)$ has only one fixed point under the lifted toric action on $\mathbb{P}^3$.*

[4] A *second canonical filtration*, that we won't use, is also defined in [Drézet 2006, §4]. The interested reader might check the given reference.

Proof. Since the $\mathbb{T}$-fixed 0-rank instantons can only be $\mathcal{O}_{l_0}(p_i)$, $i = 0, 1$, one gets two quotient maps

$$\mathcal{O}_{\mathbb{P}^3}^{\oplus 2} \twoheadrightarrow \mathcal{O}_{l_0}(p_i), \quad i = 0, 1,$$

with isomorphic kernels E_i, $i = 0, 1$. These are fixed framed instanton sheaves and all of them belong to the same class E. Hence in this case there is only one isolated point. □

The next result gives the tangent and obstruction spaces at the fixed 0 instanton pairs.

Lemma 5.11. *For the stable pairs* $\rho_i = (\mathcal{Q}_i = \mathcal{O}_{l_0}(p_i), s) \in P_1(\mathbb{P}^3, \beta = H^2)^{\mathbb{T}}$, $i = 0, 1$, *one has*

$$\mathrm{T}_{\rho_i} P_1(\mathbb{P}^3, \beta = H^2)^{\mathbb{T}} = \mathbb{C}^5, \quad \mathrm{Obs}_{\rho_i} P_1(\mathbb{P}^3, \beta = H^2)^{\mathbb{T}} = \mathbb{C}^3.$$

Proof. Recall from [Pandharipande and Thomas 2009b] that we have a triangle

$$\mathcal{Q}_i[-1] \to I^\bullet \to \mathcal{O}_{\mathbb{P}^3} \to \mathcal{Q}_i \tag{18}$$

in $D^b(\mathbb{P}^3)$, where $\mathcal{Q} = \mathcal{O}_{l_0}(p_i)$, $i = 1, 2$, and $I^\bullet := \{\mathcal{O}_{\mathbb{P}^3} \to \mathcal{Q}\}$.

Applying $\mathrm{Hom}(-, \mathcal{Q})$ on (18), one gets the sequence

$$\begin{aligned}\mathrm{Ext}^{-1}(I^\bullet, \mathcal{O}_{l_0}(p_i)) &\to \mathrm{End}(\mathcal{O}_{l_0}(p_i)) \to \mathrm{Ext}^0(\mathcal{O}_{\mathbb{P}^3}, \mathcal{O}_{l_0}(p_i)) \\ &\to \mathrm{Ext}^0(I^\bullet, \mathcal{O}_{l_0}(p_i)) \to \mathrm{Ext}^1(\mathcal{O}_{l_0}(p_i), \mathcal{O}_{l_0}(p_i)) \to \mathrm{Ext}^1(\mathcal{O}_{\mathbb{P}^3}, \mathcal{O}_{l_0}(p_i)) \\ &\to \mathrm{Ext}^1(I^\bullet, \mathcal{O}_{l_0}(p_i)) \to \mathrm{Ext}^2(\mathcal{O}_{l_0}(p_i), \mathcal{O}_{l_0}(p_i))\end{aligned} \tag{19}$$

where $\mathrm{Ext}^{-1}(I^\bullet, \mathcal{O}_{l_0}(p_i)) = 0$, as in the proof of [Pandharipande and Thomas 2009b, Lemma 1.5]. Observe that $\chi(I^\bullet, \mathcal{O}_{l_0}(p_i)) = 2$, and $\chi(\mathcal{O}_{l_0}(p_i), \mathcal{O}_{l_0}(p_i)) = 4$, by the Hirzebruch–Riemann–Roch theorem. One can easily compute that

$$\mathrm{End}(\mathcal{O}_{l_0}(p_i)) = \mathbb{C}, \quad \mathrm{Ext}^1(\mathcal{O}_{l_0}(p_i), \mathcal{O}_{l_0}(p_i)) = \mathbb{C}^4, \quad \mathrm{Ext}^2(\mathcal{O}_{l_0}(p_i), \mathcal{O}_{l_0}(p_i)) = \mathbb{C}^3,$$

and also

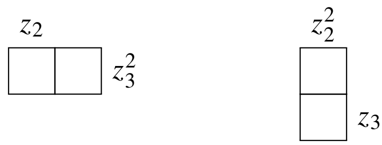
$$\mathrm{Ext}^1(\mathcal{O}_{\mathbb{P}^3}, \mathcal{O}_{l_0}(p_i)) \cong \mathrm{H}^1(\mathcal{O}_{l_0}(p_i)) = 0, \quad \mathrm{Hom}(\mathcal{O}_{\mathbb{P}^3}, \mathcal{O}_{l_0}(p_i)) \cong \mathrm{H}^0(\mathcal{O}_{l_0}(p_i)) = \mathbb{C}.$$

From (19), it follows that

$$\mathrm{Ext}^1(I^\bullet, \mathcal{O}_{l_0}(p_i)) \cong \mathrm{Ext}^2(\mathcal{O}_{l_0}(p_i), \quad \mathcal{O}_{l_0}(p_i)) \cong \mathbb{C}^3, \quad \mathrm{Ext}^0(I^\bullet, \mathcal{O}_{l_0}(p_i)) \cong \mathbb{C}^5. \ \square$$

The dimension of the tangent space at the pair $(\mathcal{Q}, s)$ is 5 as expected, since one has to choose a line in $\mathbb{P}^3$, thus a point in $\mathbb{G}(2, 4)$, and a section in $\mathbb{P}^1 = \mathbb{P}(\mathrm{H}^0(\mathcal{Q}))$. Moreover, the dimension of the obstruction space does not jump, and hence the fixed locus is, in this case, smooth.

5B2. *Stable rank* 0 *instanton pair of charge* 2*.* For a $\mathbb{T}$-fixed stable rank 0 instanton pair $(\mathcal{Q}, s)$ of charge 2, the associated Cohen–Macaulay curve $\mathcal{C}$ is a primitive double curve with ideal generated by monomials; hence one can associate to it one of the following Young diagrams:



We will only treat the case $\square\square$, the other case being very similar. The ideal sheaf of $\mathcal{I}_\mathcal{C}$, of $\mathcal{C}$, in $\mathcal{O}_{\mathbb{P}^3}$ is $\mathcal{I}_\mathcal{C} = \langle z_2, z_3^2\rangle$, and $\mathcal{C}$ is clearly a complete intersection. Moreover, it is easy to see that we have

$$0 \to \mathcal{O}_{\mathbb{P}^3}(-3) \to \mathcal{O}_{\mathbb{P}^3}(-2) \oplus \mathcal{O}_{\mathbb{P}^3}(-1) \to \mathcal{I}_\mathcal{C} \to 0,$$

with Hilbert polynomial $\chi(m) = 2m - 1$, so that $l_\mathcal{Z} = 3$. Using [Nollet 1997, Lemma 1.3] one has a sequence

$$0 \to \mathcal{I}_\mathcal{C} \to \mathcal{I}_{l_0} \to L \cong \mathcal{O}_{l_0}(-1) \to 0,$$

and hence the restriction sequence

$$0 \to \mathcal{O}_{l_0}(-1) \to \mathcal{O}_\mathcal{C} \to \mathcal{O}_{l_0} \to 0.$$

Thus the first canonical filtration of $\mathcal{O}_\mathcal{C}$ is simply $0 \subset \mathcal{O}_{l_0}(-1) \subset \mathcal{O}_\mathcal{C}$, and the graded sheaf associated to it is $\mathrm{Gr}(\mathcal{O}_\mathcal{C}) = \mathcal{O}_{l_0} \oplus \mathcal{O}_{l_0}(-1)$. This gives the generalised rank and degree, respectively, $R(\mathcal{O}_\mathcal{C}) = 2$, $\mathrm{Deg}(\mathcal{O}_\mathcal{C}) = -1$.

$\mathcal{Q}$ has first canonical filtration $0 \subset \mathcal{Q}_2 \subset \mathcal{Q}$ with a graded object $\mathrm{Gr}(\mathcal{Q}) = \mathcal{Q}|_{l_0} \oplus \mathcal{Q}_2$. Thus, one obtains the diagram

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & \mathcal{O}_{l_0}(-1) & \to & \mathcal{O}_\mathcal{C} & \to & \mathcal{O}_{l_0} & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & \mathcal{Q}_2 & \to & \mathcal{Q} & \to & \mathcal{Q}|_{l_0} & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & \mathcal{Z}_1 & \to & \mathcal{Z} & \to & \mathcal{Z}_2 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array} \tag{20}$$

Thus $\mathcal{Q}_2 = \mathcal{O}_{l_0}(-1)$, the generalised rank and degree of $\mathcal{Q}$ are, respectively, $R(\mathcal{Q}) = 2$ and $\mathrm{Deg}(\mathcal{Q}) = 2$. This leaves us with the following possibility:

Theorem 5.12. *$\mathcal{Q}|_{l_0} \cong \mathcal{O}_{l_0}(1) \oplus T$, where T is a torsion sheaf of length* 1*.*

Proof. Torsion free sheaves of generalised rank 2 on the double line are of three types [Drézet 2006, §8.2], namely line bundles, vector bundles on l_0 and the strictly torsion-free.

If $\mathcal{Q}$ is a vector bundle on l_0, then it is equal to its restriction, which contradicts the diagram (20) by the fact that $\mathcal{Q}_2 = 0$, and the snake lemma, implies that $\mathcal{Z}_1$ is a pure torsion subsheaf of $\mathcal{O}_{l_0}$.

If $\mathcal{Q}$ is a line bundle on $\mathcal{C}$, then its restriction is the line bundle $\mathcal{O}_{l_0}(3)$, which is the only possibility compatible with the right column in (20). On the other hand $\mathcal{Q}$ fits in the exact sequence

$$0 \to D \otimes \mathcal{O}_{l_0}(-1) \to \mathcal{Q} \to D \to 0,$$

where $D = \mathcal{O}_{l_0}(3)$. But this means that $\mathcal{O}_{l_0}(3) \otimes \mathcal{O}_{l_0}(-1) \cong \mathcal{O}_{l_0}(-1)$. Hence, $\mathcal{Q}$ cannot be a line bundle on $\mathcal{C}$.

Finally, in a more general situation $\mathcal{Q}$ fits in a short exact sequence

$$0 \to D \otimes \mathcal{O}_{l_0}(-1) \to \mathcal{Q} \to D \oplus T \to 0,$$

where D is a line bundle on l_0 and T is a torsion sheaf, also on l_0. Twisting the diagram by $\mathcal{O}_{\mathbb{P}^3}(-2)$ and using the vanishing conditions $\mathrm{H}^{0,1}(\mathcal{Q}(-2)) = 0$, one has that $0 < d := \deg(D) < 3$ and $\mathrm{H}^1(D(-3)) = \mathrm{H}^0(D(-2)) \oplus \mathrm{H}^0(T)$.

If $d = 2$ then $\mathrm{H}^1(D(-3)) = \mathrm{H}^1(\mathcal{O}_{l_0}(-1)) = 0\,\mathrm{H}^0(T)$. It follows that the torsion sheaf T is zero, which is not possible, as the restriction $\mathcal{Q}|_{l_0}$ is not locally free.

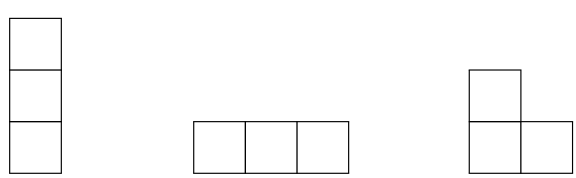
If $d = 1$ then $\mathrm{H}^1(D(-3)) = \mathrm{H}^1(\mathcal{O}_{l_0}) = \mathbb{C} = \mathrm{H}^0(T)$. Thus T is a torsion sheaf of length 1, and it follows, from (20), that $\mathcal{Z}_2$ is the structure sheaf of 2 points. □

Corollary 5.13. $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(2)$ *has at least* 2 *irreducible component.*

Proof. From Theorem 5.6, and as we saw in the beginning of this section, there are two possible Young diagrams for the support. To each one of these curves there is one possible filtration $0 \subset \mathcal{O}_{l_0} \subset \mathcal{Q}$ with restriction $\mathcal{O}_{l_0}(1) \oplus T$, where $\mathrm{length}(T) = 1$. Finally, by Theorem 5.12 and the fact that $\mathcal{Q}$ might have nontrivial deformation in each case, the result follow. □

We remark that the double curve $\mathcal{C}$ can be deformed into two lines intersecting in a point, but we don't know if one can deform the 0-rank instanton sheaves $\mathcal{Q}$ into torsion free sheaves on the reducible curve formed by two intersecting lines. This is a hard problem and should be investigated in the future.

5B3. *Stable rank* 0 *instanton pair of charge* 3. In this case one has 3 possible associated Young diagrams, namely



In the following, we will treat only the primitive cases, so we can concentrate on the horizontal Young diagram, the vertical case being similar. The canonical filtration of the triple curve $\mathcal{C}$ is given by

$$0 \subset \mathcal{L}_3 = \mathcal{O}_{l_0}(-1) \subset \mathcal{L}_2 \subset \mathcal{O}_{\mathcal{C}}$$

with quotients

$$\mathcal{O}_{\mathcal{C}}/\mathcal{L}_2 = \mathcal{O}_{l_0}, \quad \mathcal{O}_{\mathcal{C}}/\mathcal{L}_3 = \mathcal{O}_{\mathcal{C}_2} \quad \text{and} \quad \mathcal{L}_2/\mathcal{L}_3 = \mathcal{O}_{l_0}(-1).$$

We also recall that $\chi(\mathcal{O}_{\mathcal{C}}(m)) = 3m$ and $\chi(\mathcal{Q}(m)) = 3m+6$. On the other hand one has a canonical filtration $0 \subset \mathcal{Q}_3 \subset \mathcal{Q}_2 \subset \mathcal{Q}$. Twisting by $\mathcal{O}_{\mathbb{P}^3}(-2)$ and using the instanton conditions $\mathrm{H}^{0,1}(\mathcal{Q}(-2)) = 0$, one has

$$\begin{aligned}
\mathrm{H}^0(\mathcal{Q}_2(-2)) &= \mathrm{H}^0(\mathcal{Q}_3(-2)) = 0,\\
\mathrm{H}^1(\mathcal{Q}|_{l_0}(-2)) &= \mathrm{H}^1(\mathcal{Q}|_{\mathcal{C}_2}(-2)) = 0,\\
\mathrm{H}^1(\mathcal{Q}_2(-2)) &= \mathrm{H}^0(\mathcal{Q}|_{l_0}(-2)),\\
\mathrm{H}^1(\mathcal{Q}_3(-2)) &= \mathrm{H}^0(\mathcal{Q}|_{\mathcal{C}_2}(-2)).
\end{aligned} \tag{21}$$

Moreover, $\mathcal{Q}_2(-2)$ is a generalised rank 2 sheaf on $\mathcal{C}_2$ and $\mathcal{Q}_3(-2)$ is a line bundle on l_0. Then we have the following exact sequence, associated to the canonical filtration of $\mathcal{Q}_2(-2)$,

$$0 \to D \otimes \mathcal{O}_{l_0}(-1) \to \mathcal{Q}_2(-2) \to D \oplus T_2 \to 0,$$

in which $D = \mathcal{Q}_3(-1)$ and T_2 is pure torsion sheaf on l_0. But from the first two equations of (21), the degree d, of $D = \mathcal{Q}_3(-1)$, satisfies $-2 < d < 1$, i.e., $\deg(\mathcal{Q}_3) = 0$, or $\deg(\mathcal{Q}_3) = 1$.

The next step is to consider the commutative diagrams

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{L}_2(-2) & \longrightarrow & \mathcal{O}_{\mathcal{C}}(-2) & \longrightarrow & \mathcal{O}_{l_0}(-2) & \longrightarrow & 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{Q}_2(-2) & \longrightarrow & \mathcal{Q}(-2) & \longrightarrow & \mathcal{Q}|_{l_0}(-2) & \longrightarrow & 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \overline{\mathcal{Z}} & \longrightarrow & \mathcal{Z}_6 & \longrightarrow & \widetilde{\mathcal{Z}} & \longrightarrow & 0\\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array} \tag{22}$$

and

$$
\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{O}_{l_0}(-4) & \longrightarrow & \mathcal{L}_2(-2) & \longrightarrow & \mathcal{O}_{l_0}(-3) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{Q}_3(-2) & \longrightarrow & \mathcal{Q}_2(-2) & \longrightarrow & \mathcal{Q}_3(-1) \oplus T_2 & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{Z}_2 & \longrightarrow & \overline{\mathcal{Z}} & \longrightarrow & \mathcal{Z}_2 \oplus T_2 & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}
\tag{23}
$$

associated to the sequence

$$0 \to \mathcal{O}_{\mathcal{C}} \to \mathcal{Q} \to \mathcal{Z}_6 \to 0$$

and the canonical filtrations of $\mathcal{O}_{\mathcal{C}}$ and $\mathcal{Q}$. Here we remind the reader that $\mathcal{Z}_6$ has length 6 and $\mathcal{Z}_2$ is a torsion sheaf of length 2, if $\mathcal{Q}_3 = \mathcal{O}_{l_0}$, or length 3, if $\mathcal{Q}_3 = \mathcal{O}_{l_0}(1)$, as it clearly appears from left column of diagram (23).

Suppose that $\mathcal{Q}_3 = \mathcal{O}_{l_0}(1)$. Then, from the middle row of diagram (23), we have $\chi(\mathcal{Q}_2(-2)) = t_2 + 1$. In particular, $\chi(\mathcal{Q}_2(-2)) > 0$, since $t_2 \geq 0$. On the other hand, we have $\mathrm{H}^0(\mathcal{Q}_2(-2)) = 0$, and it follows that $\chi(\mathcal{Q}_2(-2)) = -\dim \mathrm{H}^1(\mathcal{Q}_2(-2)) < 0$. Thus $\mathcal{Q}_3$ cannot be $\mathcal{O}_{l_0}(1)$, and we are left with $\mathcal{Q}_3 = \mathcal{O}_{l_0}$. In this case we have $\chi(\mathcal{Q}_2(-2)) = t_2 - 1$ and t_2 must be 0 or 1, since $\chi(\mathcal{Q}_2(-2)) \leq 0$.

If $t_2 = 0$, then $T_2 = 0$, $\mathcal{Q}_2(-2) = \mathcal{L}_2$, since the middle row of (23) is exactly the restriction sequence, given by the canonical filtration, of $\mathcal{L}_2$ to l_0. Furthermore, from the lower rows of (22) and (23). It follows that $\overline{\mathcal{Z}}$ has length 4, thus $\widetilde{\mathcal{Z}}$ has length 2. By using the right column of (22) one has $\chi(\mathcal{Q}|_{l_0}(-2)) = 1$, Hence $\mathcal{Q}|_{l_0} = \mathcal{O}_{l_0}$.

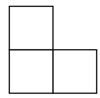
Finally if $t_2 = 1$ then, from the middle row of (23), one can see that $\mathcal{Q}_2$ also satisfies $\mathrm{H}^{0,1}(\mathcal{Q}_2(-2)) = 0$; hence it is an rank 0 instanton sheaf over $\mathcal{C}_2$. Moreover, the length of $\overline{\mathcal{Z}}$ is equal to 5 and it follows, again, from the right column of diagram (23) that $\widetilde{\mathcal{Z}}$ has length 1 and $\mathcal{Q}|_{l_0} = \mathcal{O}_{l_0}(1)$. This proves the following:

Theorem 5.14. *For $c = 3$ and $\mathcal{C}$ is primitive and monomial, the sheaf $\mathcal{Q}$ has a canonical filtration $0 \subset \mathcal{Q}_3 = \mathcal{O}_{l_0} \subset \mathcal{Q}_2 \subset \mathcal{Q}$ in which*

(i) $\mathcal{Q}_2 = \mathcal{L}_2(2)$, $\mathcal{Q}|_{l_0} = \mathcal{O}_{l_0}(2)$, *and* $\mathcal{Q}_2/\mathcal{Q}_3 = \mathcal{O}_{l_0}(1)$, *or*

(ii) $\mathcal{Q}_2$ *is a rank* 0 *instanton sheaf on* $\mathcal{C}_2 \subset \mathcal{C}$, $\mathcal{Q}|_{l_0} = \mathcal{O}_{l_0}(1)$, *and* $\mathcal{Q}_2/\mathcal{Q}_3 = \mathcal{O}_{l_0}(1) \oplus T_2$, *where* T_2 *is a torsion sheaf of length* 1.

Now we turn our attention to the first nonprimitive case, that is, when the corresponding Young diagram is

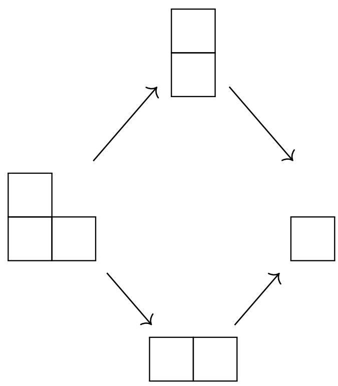


This is the case of the (affine) ideal $\langle x, y\rangle^2$. It is easy to check that the restriction map is given by

$$0 \to \mathcal{O}_{l_0}(-1)^{\oplus 2} \to \mathcal{O}_C \to \mathcal{O}_{l_0} \to 0.$$

There are two filtrations represented as below:

(24)



Note that the double structures in the middle column are primitive, and although there is no unique canonical filtration, we still manage to compute the resulting possible pure sheaves $\mathcal{Q}$. We use for instance the filtration given by the upper arrow of (24), and as in the previous theorem, we consider restriction diagrams as (22) and (23):

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{O}_{l_0}(-3)^{\oplus 2} & \longrightarrow & \mathcal{O}_C(-2) & \longrightarrow & \mathcal{O}_{l_0}(-2) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{Q}_2(-2) & \longrightarrow & \mathcal{Q}(-2) & \longrightarrow & \mathcal{Q}|_{l_0}(-2) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \overline{\mathcal{Z}} & \longrightarrow & \mathcal{Z}_5 & \longrightarrow & \widetilde{\mathcal{Z}} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array} \tag{25}$$

and

$$
\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{O}_{l_0}(-3) & \longrightarrow & \mathcal{O}_{\mathcal{C}}(-2) & \longrightarrow & \mathcal{O}_{\mathcal{C}_2}(-2) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathcal{Q}_3(-2) & \longrightarrow & \mathcal{Q}(-2) & \longrightarrow & \mathcal{Q}|_{\mathcal{C}_2}(-2) & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \hat{\mathcal{Z}}_2 & \longrightarrow & \mathcal{Z}_5 & \longrightarrow & \check{\mathcal{Z}}_2 & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}
\tag{26}
$$

We recall that $\chi(\mathcal{Q}(m)) = 3m+6$, $\chi(\mathcal{O}_{\mathcal{C}}(m)) = 3m+1$, so that $\mathcal{Z}_5$ has length 5.

Theorem 5.15. *For $c = 3$ and $\mathcal{C}$ the nonprimitive monomial curve, the sheaf $\mathcal{Q}$ has a filtration $0 \subset \mathcal{O}_3 \subset \mathcal{Q}_2 \subset \mathcal{Q}$ such that*

(i) $\mathcal{Q}|_{l_0} = \mathcal{O}_{l_0}(1) \oplus T_3$, *where T_3 is a torsion sheaf of length 3, $\mathcal{Q}_2$ is a sheaf on $\mathcal{C}_2$ with restriction sequence*

$$0 \to \mathcal{O}_{l_0}(-1) \to \mathcal{Q}_2 \to \mathcal{O}_{l_0} \to 0,$$

and $\mathcal{Q}|_{\mathcal{C}_2}$ is also a sheaf on $\mathcal{C}_2$ with restriction sequence

$$0 \to \mathcal{O}_{l_0} \to \mathcal{Q}|_{\mathcal{C}_2} \to \mathcal{O}_{l_0}(-1) \oplus T_3 \to 0, \quad \textit{or}$$

(ii) $\mathcal{Q}|_{l_0} = \mathcal{O}_{l_0}(2) \oplus T_1$, *where T_1 is a torsion sheaf of length 1, $\mathcal{Q}_2$ is a sheaf on $\mathcal{C}_2$ with restriction sequence*

$$0 \to \mathcal{O}_{l_0}(-1) \to \mathcal{Q}_2 \to \mathcal{O}_{l_0} \oplus T_1 \to 0,$$

and $\mathcal{Q}|_{\mathcal{C}_2}$ is also a sheaf on $\mathcal{C}_2$ with restriction sequence

$$0 \to \mathcal{O}_{l_0}(1) \to \mathcal{Q}|_{\mathcal{C}_2} \to \mathcal{O}_{l_0}(2) \oplus T_1 \to 0, \quad \textit{or}$$

(iii) $\mathcal{Q}|_{l_0} = \mathcal{O}_{l_0}(2)$, *$\mathcal{Q}_2$ is a sheaf on $\mathcal{C}_2$ with restriction sequence*

$$0 \to \mathcal{O}_{l_0} \to \mathcal{Q}_2 \to \mathcal{O}_{l_0}(1) \to 0,$$

and

$$0 \to \mathcal{O}_{l_0}(1) \to \mathcal{Q}|_{\mathcal{C}_2} \to \mathcal{O}_{l_0}(2) \to 0,$$

Proof. The proof strategy is similar to that of Theorem 5.14, by arguing on the length $\tilde{z}$, of $\widetilde{\mathcal{Z}}$ in (25); first, note that $0 \leq \tilde{z} \leq 5$. If $\tilde{z} = 0$, then $\mathcal{Q}_2(-2) = \mathcal{O}_{l_0}(-3)^{\oplus 2}$ and $\chi(\mathcal{Q}|_{l_0}(-2)) = 4$. Moreover, $\mathcal{Q}|_{l_0}(-2) = \mathcal{L} \oplus T$. By putting $d = \deg(\mathcal{L})$ and

$t = \text{length}(T)$, one has $d+t=3$. On the other hand, $\mathcal{Q}_3(-2)=\mathcal{O}_{l_0}(-3)$, and from the middle row of (26), $\chi(\mathcal{Q}|_{C_2}(-2))=2$. Furthermore, $\mathcal{Q}|_{C_2}(-2)$ fits into the restriction sequence

$$0 \to \mathcal{L}(-1) \to \mathcal{Q}|_{C_2}(-2) \to \mathcal{L} \oplus T \to 0, \tag{27}$$

since $(\mathcal{Q}|_{C_2})|_{l_0} = \mathcal{Q}|_{l_0}$. But this implies that $\mathcal{L}=\mathcal{O}_{l_0}(-2)$ and $\text{length}(T)=5$, which leads to $\mathrm{H}^1(\mathcal{Q}_2(-2))=\mathbb{C}^4$ and $\mathrm{H}^0(\mathcal{Q}|_{l_0}(-2))=\mathbb{C}^5$, contradicting the third equation of (21). Thus $\tilde{z}\neq 0$. In the same fashion, one proves that $\tilde{z}$ cannot be equal to 4, nor 5.

For the rest of the cases $\tilde{z}=1, 2, 3$, one can first write

$$0 \to \mathcal{L}_2(-1) \to \mathcal{Q}_2(-2) \to \mathcal{L}_2 \oplus T \to 0,$$

from which one has $\mathcal{Q}_3=\mathcal{L}_2(-1)$ and $\chi(\mathcal{Q}_2(-2))=2d_2+t+1$ and set $t=\text{length}(T)$ and $d_2=\deg\mathcal{L}_2$. Then, by using the left column of (25), one has the following table:

(28)

$\tilde{z}$	$\chi(\mathcal{Q}_2(-2))$	$\bar{z}$	$\chi(\mathcal{Q}\|_{l_0}(-2))$
1	-3	4	3
2	-2	3	2
3	-1	2	1

Recall that $\bar{z}$ is the length of $\overline{\mathcal{Z}}$ in (25). In what follows we analyse the case in the first row of (28). The other cases can be treated similarly.

When $\tilde{z}=1$ one has $2l_2+t+1=-3$. Since the length $t\geq 0$, one has $d_2=-2$ and $t=0$ or $d_2=-3$ and $t=2$ or $d_2=-4$ and $t=4$. However, the last two cases cannot hold since $\mathcal{Q}_3(-2)=\mathcal{L}_2(-1)$ would have degree less than -3, contradicting the first row of (26). Hence we end up with $d_2=-2$, $t=0$ and $\mathcal{Q}_3(-2)=\mathcal{O}_{l_0}(-3)$. Now, writing $\mathcal{Q}|_{C_2}(-2)$ as in (27) one should have $2d+t=1$. Again, by using the fact that $\mathcal{Q}|_{l_0}(-2)=\mathcal{L}\oplus T$, it turns out that the only possibility is $d=-1$ and $t=3$. Hence

$$0 \to \mathcal{O}_{l_0}(-2) \to \mathcal{Q}|_{C_2}(-2) \to \mathcal{O}_{l_0}(-1)\oplus T_3 \to 0,$$
$$0 \to \mathcal{O}_{l_0}(-3) \to \mathcal{Q}_2(-2) \to \mathcal{O}_{l_0}(-2) \to 0.$$ □

Remark 5.16. (I) When $\mathcal{C}$ is primitive, the graded object $\mathrm{Gr}(\mathcal{Q})$, associated to the canonical filtration of $\mathcal{Q}$, can be computed from Theorem 5.14; in case (i) one has

$$\mathrm{Gr}(\mathcal{Q})=\mathcal{O}_{l_0}(2)\oplus\mathcal{O}_{l_0}(1)\oplus\mathcal{O}_{l_0},$$

hence $\mathcal{Q}$ is a generalised rank 3 quasi locally free sheaf on $\mathcal{C}$ [Drézet 2006, Corollary 5.1.4]. In case (ii), one has

$$\mathrm{Gr}(\mathcal{Q})=\mathcal{O}_{l_0}(1)^{\oplus 2}\oplus\mathcal{O}_{l_0}\oplus T_2,$$

and T_2 has length 1. Hence $\mathcal{Q}$ is, in this case, a generalised rank 3 sheaf on $\mathcal{C}$.

(II) For the nonprimitive case, one can also compute the graded object, with respect to the chosen filtration, from Theorem 5.15, namely

(i) $\mathrm{Gr}(\mathcal{Q}) = \mathcal{O}_{l_0}(-1) \oplus \mathcal{O}_{l_0} \oplus \mathcal{O}_{l_0}(1) \oplus T_3$, and T_3 has length 3, or

(ii) $\mathrm{Gr}(\mathcal{Q}) = \mathcal{O}_{l_0}(2) \oplus \mathcal{O}_{l_0} \oplus \mathcal{O}_{l_0}(-1) \oplus T_1^{\oplus 2}$, and T_1 has length 1, or

(iii) $\mathrm{Gr}(\mathcal{Q}) = \mathcal{O}_{l_0}(2) \oplus \mathcal{O}_{l_0}(1) \oplus \mathcal{O}_{l_0}$.

(III) Theorems 5.14 and 5.15 show that there are at least 7 components in $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(3)$; 3 nonprimitive cases and 4 primitive cases, counting both the horizontal and the vertical Young diagrams.

For a given integer m, we now denote by $\mathrm{p}(m)$ the number of its (2-dimensional) partitions. We recall that $l_{\mathcal{Z}}(\nu(c))$ denotes the length of $\mathcal{Z}$, for the multiple structure associated to a partition $\nu(c)$ of c.

If we consider the whole set of monomial multiple structures, not only the primitive ones, then we get the following:

Lemma 5.17. *The fixed locus $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c)$ splits as a union*

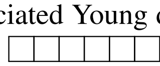
$$\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c) = \bigcup_{\nu(\boldsymbol{c})} \mathcal{M}(\nu(\boldsymbol{c})).$$

Thus the least number of such irreducible components in $\mathcal{M}^{\mathbb{T}}_{\mathbb{P}^3}(c)$ is given by the number $\mathrm{p}(c)$, *of partitions $\nu(c)$, of c, and can be expressed generating function*

$$\sum_{c=0}^{\infty} \mathrm{p}(c) x^c = \prod_{k=1}^{\infty} \frac{1}{1-x^k}.$$

Proof. This is obtained by enumerating the possible Young diagrams, hence enumerating partitions $\nu(c)$ of c. □

We remark that these are completely disconnected components. But as seen, in the case $c = 3$, there might be more than one component for the same partition.

5B4. *Stable rank* 0 *instanton pair of charge c with primitive support.* We now describe the case in which the support is a primitive multiple line. For the pair $(\mathcal{Q}, s)$ of charge c, the associated Cohen–Macaulay curve $\mathcal{C}$ is a primitive multiple curve with ideal whom associated Young diagram is a column or a line. As in the last section we treat the case .

This time we have $\mathcal{I}_{\mathcal{C}} = \langle z_2, z_3^c \rangle$ for which $\mathcal{C}$ is a complete intersection. Its resolution is

$$0 \to \mathcal{O}_{\mathbb{P}^3}(-c-1) \to \mathcal{O}_{\mathbb{P}^3}(-c) \oplus \mathcal{O}_{\mathbb{P}^3}(-1) \to \mathcal{I}_{\mathcal{C}} \to 0,$$

with Hilbert polynomial $\chi(\mathcal{O}_{\mathcal{C}}(m)) = cm - \frac{1}{2}c(c-3)$, and length $l_{\mathcal{Z}} = \frac{1}{2}c(c+1)$.

The canonical filtration of supports is represented by

$$\begin{array}{ccccccccc}
\mathcal{I}_{l_0} & \supset & \mathcal{I}_{\mathcal{C}_2} & \supset & \cdots & \supset & \mathcal{I}_{\mathcal{C}_{c-1}} & \supset & \mathcal{I}_{\mathcal{C}} \\
\square & \subset & \square\square & \subset & \cdots & \subset & \square\square\square\square\square & \subset & \square\square\square\square\square\square \\
l_0 & \subset & \mathcal{C}_2 & \subset & \cdots & \subset & \mathcal{C}_{c-1} & \subset & \mathcal{C}
\end{array}$$

and we have sequences

$$0 \to \mathcal{I}_{\mathcal{C}_2} \to \mathcal{I}_{l_0} \to L \cong \mathcal{O}_{l_0}(-1) \to 0,$$
$$0 \to \mathcal{I}_{\mathcal{C}_{i+1}} \to \mathcal{I}_{\mathcal{C}_i} \to L \cong \mathcal{O}_{l_0}(-1)^{\otimes i} \to 0,$$

and hence restriction sequences

$$0 \to \mathcal{O}_{l_0}(-1) \to \mathcal{O}_{\mathcal{C}} \to \mathcal{O}_{l_0} \to 0,$$
$$0 \to \mathcal{O}_{l_0}(-i) \to \mathcal{O}_{\mathcal{C}_{i+1}} \to \mathcal{O}_{\mathcal{C}_i} \to 0$$

for $1 \leq i \leq c-1$.

On the other hand, the first canonical filtration of $\mathcal{O}_{\mathcal{C}}$ reads as

$$\mathcal{L}_{c+1} = 0 \subset \mathcal{L}_c \subset \cdots \subset \mathcal{L}_2 \subset \mathcal{O}_{\mathcal{C}},$$

where $\mathcal{O}_{\mathcal{C}}/\mathcal{L}_{i+1} \cong \mathcal{O}_{\mathcal{C}_i}$.

Lemma 5.18. *The graded sheaf, the generalised degree and the generalised rank of $\mathcal{O}_{\mathcal{C}}$ are given, respectively, by*

$$\mathrm{Gr}(\mathcal{O}_{\mathcal{C}}) = \bigoplus_{i=0}^{c-1} \mathcal{O}_{l_0}(-i), \quad \mathrm{Deg}(\mathcal{O}_{\mathcal{C}}) = -\frac{c(c-1)}{2} \quad \textit{and} \quad R(\mathcal{O}_{\mathcal{C}}) = c.$$

Proof. By using diagrams

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & \mathcal{L}_{i+1} & = & \mathcal{L}_{i+1} & & & & \\
 & & \downarrow & & \downarrow & & & & \\
0 & \longrightarrow & \mathcal{L}_i & \longrightarrow & \mathcal{O}_{\mathcal{C}} & \longrightarrow & \mathcal{O}_{\mathcal{C}_{i-1}} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \| & & \\
0 & \longrightarrow & \mathcal{O}_{l_0}(-i+1) & \longrightarrow & \mathcal{O}_{\mathcal{C}_i} & \longrightarrow & \mathcal{O}_{\mathcal{C}_{i-1}} & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

one gets the graded sheaf. The generalised degree and rank follow easily by applying their definitions. $\square$

By Theorem 5.9 one gets $\chi(\mathcal{O}_{\mathcal{C}}(m)) = cm - \frac{1}{2}c(c-3)$. Thus the graded sheaf associated to $\mathcal{O}_{\mathcal{C}}$ is $\mathrm{Gr}(\mathcal{O}_{\mathcal{C}}) = \bigoplus_{i=0}^{c-1} \mathcal{O}_{l_0}(-i)$, and the generalised rank and degree are, respectively, $R(\mathcal{O}_{\mathcal{C}}) = c$, $\mathrm{Deg}(\mathcal{O}_{\mathcal{C}}) = -\frac{1}{2}c(c-1)$.

We remark that we do not know whether the above fixed components intersect the closure of the framed locally free instanton moduli, in general. We think that this problem is related to *reachability* of sheaves, on multiple structure [Drézet 2017]. Nevertheless, for charge $c = 1$, the answer is positive; the sheaf $\ker(\mathcal{O}_{\mathbb{P}^3}^{\oplus 2} \to \mathcal{Q})$, in Section 5B1, is in the closure of the moduli of locally free framed instanton bundles [Jardim et al. 2017b, §6]. Furthermore, if $c = 2$ one can deform the (monomial) double curve into a union of two curves intersecting at a point. Moreover, the moduli space of instantons of charge $c = 2$ is irreducible as proved in [Jardim et al. 2017a, Proposition 7]. Thus the 0 instanton sheaf should deform, from sheaf on the double curve, to a sheaf on the reduced curve. Hence, the fixed component is in the closure of the moduli space of framed instantons. For higher values of the charge this is a difficult problem to answer. Since we think this is true, we close this note by writing:

Conjecture. *The fixed components, under the lifted toric action on* $\mathbb{P}^3$, *intersect the closure of the locally free component in the moduli space of framed instantons.*

Acknowledgments

I would like to thank Marcos Jardim for the useful discussions we had during my few short visits to IMECC-UNICAMP and his valuable remarks about the first draft. I am also thankful to the referees for their corrections and suggestions.

References

[Atiyah et al. 1978] M. F. Atiyah, N. J. Hitchin, V. G. Drinfeld, and Y. I. Manin, "Construction of instantons", *Phys. Lett. A* **65**:3 (1978), 185–187. MR Zbl

[Baranovsky 2000] V. Baranovsky, "Moduli of sheaves on surfaces and action of the oscillator algebra", *J. Differential Geom.* **55**:2 (2000), 193–227. MR Zbl

[Belavin et al. 1975] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Y. S. Tyupkin, "Pseudoparticle solutions of the Yang–Mills equations", *Phys. Lett. B* **59**:1 (1975), 85–87. MR

[Brion and Peyre 2002] M. Brion and E. Peyre, "The virtual Poincaré polynomials of homogeneous spaces", *Compos. Math.* **134**:3 (2002), 319–335. MR Zbl

[Bruzzo et al. 2011] U. Bruzzo, R. Poghossian, and A. Tanzini, "Poincaré polynomial of moduli spaces of framed sheaves on (stacky) Hirzebruch surfaces", *Comm. Math. Phys.* **304**:2 (2011), 395–409. MR Zbl

[Donaldson 1984] S. K. Donaldson, "Instantons and geometric invariant theory", *Comm. Math. Phys.* **93**:4 (1984), 453–460. MR Zbl

[Donaldson and Kronheimer 1990] S. K. Donaldson and P. B. Kronheimer, *The geometry of four-manifolds*, Oxford Univ. Press, 1990. MR Zbl

[Drézet 2006] J.-M. Drézet, "Faisceaux cohérents sur les courbes multiples", *Collect. Math.* **57**:2 (2006), 121–171. MR Zbl

[Drézet 2009] J.-M. Drézet, "Faisceaux sans torsion et faisceaux quasi localement libres sur les courbes multiples primitives", *Math. Nachr.* **282**:7 (2009), 919–952. MR Zbl

[Drézet 2017] J.-M. Drézet, "Reachable sheaves on ribbons and deformations of moduli spaces of sheaves", *Int. J. Math.* **28**:12 (2017), art. id. 1750086. MR Zbl

[Ein 1982] L. Ein, "Some stable vector bundles on $\mathbb{P}^4$ and $\mathbb{P}^5$", *J. Reine Angew. Math.* **337** (1982), 142–153. MR Zbl

[Frenkel and Jardim 2008] I. B. Frenkel and M. Jardim, "Complex ADHM equations and sheaves on $\mathbb{P}^3$", *J. Algebra* **319**:7 (2008), 2913–2937. MR Zbl

[Gargate and Jardim 2016] M. Gargate and M. Jardim, "Singular loci of instanton sheaves on projective space", *Int. J. Math.* **27**:7 (2016), art. id. 1640006. MR Zbl

[Grothendieck 1958] A. Grothendieck, "Torsion homologique et sections rationelles", exposé 5 in *Anneaux de Chow et applications*, Séminaire Claude Chevalley **3**, Sec. Math., Paris, 1958.

[Hartshorne 1980] R. Hartshorne, "Stable reflexive sheaves", *Math. Ann.* **254**:2 (1980), 121–176. MR Zbl

[Hauzer and Langer 2011] M. Hauzer and A. Langer, "Moduli spaces of framed perverse instantons on $\mathbb{P}^3$", *Glasg. Math. J.* **53**:1 (2011), 51–96. MR Zbl

[Henni et al. 2015] A. A. Henni, M. Jardim, and R. V. Martins, "ADHM construction of perverse instanton sheaves", *Glasg. Math. J.* **57**:2 (2015), 285–321. MR Zbl

[Horrocks 1964] G. Horrocks, "Vector bundles on the punctured spectrum of a local ring", *Proc. Lond. Math. Soc.* (3) **14**:4 (1964), 689–713. MR Zbl

[Huybrechts and Lehn 1997] D. Huybrechts and M. Lehn, *The geometry of moduli spaces of sheaves*, Aspects Math. **E31**, Vieweg & Sohn, Braunschweig, Germany, 1997. MR Zbl

[Jardim 2006] M. Jardim, "Instanton sheaves on complex projective spaces", *Collect. Math.* **57**:1 (2006), 69–91. MR Zbl

[Jardim and Verbitsky 2014] M. Jardim and M. Verbitsky, "Trihyperkähler reduction and instanton bundles on $\mathbb{CP}^3$", *Compos. Math.* **150**:11 (2014), 1836–1868. MR Zbl

[Jardim et al. 2017a] M. Jardim, M. Maican, and A. S. Tikhomirov, "Moduli spaces of rank 2 instanton sheaves on the projective space", *Pacific J. Math.* **291**:2 (2017), 399–424. MR Zbl

[Jardim et al. 2017b] M. Jardim, D. Markushevich, and A. S. Tikhomirov, "Two infinite series of moduli spaces of rank 2 sheaves on $\mathbb{P}^3$", *Ann. Mat. Pura Appl.* (4) **196**:4 (2017), 1573–1608. MR Zbl

[Jardim et al. 2018] M. Jardim, D. Markushevich, and A. S. Tikhomirov, "New divisors in the boundary of the instanton moduli space", *Mosc. Math. J.* **18**:1 (2018), 117–148. MR Zbl

[Le Potier 1995] J. Le Potier, "Faisceaux semi-stables et systèmes cohérents", pp. 179–239 in *Vector bundles in algebraic geometry* (Durham, UK, 1993), edited by N. J. Hitchin et al., Lond. Math. Soc. Lect. Note Ser. **208**, Cambridge Univ. Press, 1995. MR Zbl

[Nakajima 1994] H. Nakajima, "Resolutions of moduli spaces of ideal instantons on $\mathbb{R}^4$", pp. 129–136 in *Topology, geometry and field theory*, edited by K. Fukaya et al., World Sci., River Edge, NJ, 1994. MR Zbl

[Nakajima 1999] H. Nakajima, *Lectures on Hilbert schemes of points on surfaces*, Univ. Lect. Ser. **18**, Amer. Math. Soc., Providence, RI, 1999. MR Zbl

[Nakajima 2011] H. Nakajima, "Quiver varieties and cluster algebras", *Kyoto J. Math.* **51**:1 (2011), 71–126. MR Zbl

[Nakajima and Yoshioka 2005] H. Nakajima and K. Yoshioka, "Instanton counting on blowup, I: 4-dimensional pure gauge theory", *Invent. Math.* **162**:2 (2005), 313–355. MR Zbl

[Nollet 1997] S. Nollet, "The Hilbert schemes of degree three curves", *Ann. Sci. École Norm. Sup.* (4) **30**:3 (1997), 367–384. MR Zbl

[Pandharipande and Thomas 2009a] R. Pandharipande and R. P. Thomas, "The 3-fold vertex via stable pairs", *Geom. Topol.* **13**:4 (2009), 1835–1876. MR Zbl

[Pandharipande and Thomas 2009b] R. Pandharipande and R. P. Thomas, "Curve counting via stable pairs in the derived category", *Invent. Math.* **178**:2 (2009), 407–447. MR Zbl

[Serre 1958] J.-P. Serre, "Espaces fibrés algébriques", exposé 1 in *Anneaux de Chow et applications*, Séminaire Claude Chevalley **3**, Sec. Math., Paris, 1958.

[Tikhomirov 2012] A. S. Tikhomirov, "Moduli of mathematical instanton vector bundles with odd c_2 on projective space", *Izv. Ross. Akad. Nauk Ser. Mat.* **76**:5 (2012), 143–224. In Russian; translated in *Izv. Math.* **76**:5 (2012), 991–1073. MR Zbl

[Tikhomirov 2013] A. S. Tikhomirov, "Moduli of mathematical instanton vector bundles with even c_2 on projective space", *Izv. Ross. Akad. Nauk Ser. Mat.* **77**:6 (2013), 139–168. In Russian; translated in *Izv. Math.* **77**:6 (2013), 1195–1223. MR Zbl

[Vatne 2002] J. E. Vatne, "Multiple structures", preprint, 2002. arXiv

[Vatne 2012] J. E. Vatne, "Monomial multiple structures", *Ann. Univ. Ferrara Sez. VII Sci. Mat.* **58**:1 (2012), 199–215. MR Zbl

Received August 17, 2018. Revised October 10, 2019.

ABDELMOUBINE AMAR HENNI
DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE FEDERAL DE SANTA CATARINA
FLORIANÓPOLIS, SC
BRAZIL

henni.amar@ufsc.br

PACIFIC JOURNAL OF MATHEMATICS
Vol. 308, No. 1, 2020
https://doi.org/10.2140/pjm.2020.308.73

THE AZIMUTHAL EQUIDISTANT PROJECTION FOR A FINSLER MANIFOLD BY THE EXPONENTIAL MAP

Nobuhiro Innami, Yoe Itokawa, Toshiki Kondo, Tetsuya Nagano and Katsuhiro Shiohama

Let (M, F) be a geodesically forward complete Finsler manifold and $p \in M$. We observe how the preimage of a curve in M under exponential map at p can behave in the tangent space T_pM at p, when the curve approaches a conjugate cut point of p without crossing the cut locus of p. After this investigation, we may regard the internal region of a tangent cut locus of $p \in M$ as the development of M. We deal with isometry problems of Finsler manifolds and differentiability conditions of cut loci.

1. Introduction

Let (M, F) be a geodesically forward complete Finsler manifold without boundary and $\exp_p : T_pM \to M$ the exponential map at $p \in M$. Then $\exp_p$ is C^∞ on $T_pM \setminus \{0\}$ and C^1 at $0 \in T_pM$ (see [Shen 2001, Theorem 11.1.1]). Let $d(p, q)$ denote the distance from p to q induced by F and $S_pM := \{v \mid v \in T_pM \text{ with } F(p, v) = 1\}$.

For a tangent vector $v \in S_pM$ we define numbers $\rho(v), \lambda(v) \in (0, \infty]$ as follows:

$$\rho(v) = \sup\{s > 0 \mid d(p, \exp_p(tv)) = t \text{ for any } t \in (0, s)\},$$
$$\lambda(v) = \sup\{s > 0 \mid d\exp_p|_{tv} \text{ is nonsingular for any } t \in (0, s)\}.$$

It is well known that ρ and λ are continuous on the domain in S_pM where they are bounded. Let $\nu(v)$ denote the dimension of the kernel $N(v)$ of $d\exp_p|_{\lambda(v)v}$. It follows from the implicit function theorem that if ν is constant in an open set $U \subset S_pM$, then λ is C^∞ on U. In particular, if $\dim M = 2$, then λ is C^∞ on the domain U in S_pM where λ is bounded.

We call $\widetilde{C}(p) := \{\rho(v)v \mid v \in S_pM\}$ the *tangent cut locus* of p, and $C(p) := \exp_p(\widetilde{C}(p))$ the *cut locus* of p. In a similar way we define the *first tangent conjugate locus* $\widetilde{Q}(p) := \{\lambda(v)v \mid v \in S_pM\}$ and the *first conjugate locus* $Q(p) := \exp_p(\widetilde{Q}(p))$ of p. We call a point $q \in \exp_p(\widetilde{C}(p) \cap \widetilde{Q}(p))$ a *conjugate cut point* of p. We say

---

The research of Shiohama was partially supported by JSPS KAKENHI grant number 18K03314.
MSC2010: primary 53C20; secondary 53C22.
Keywords: Finsler manifold, cut locus, azimuthal equidistant projection, exponential map.

that a point $q \in C(p)$ is a *nonconjugate cut point* of p if q is not a conjugate cut point of p.

Weinstein [1968] has proved that any compact differentiable manifold M not homeomorphic to the 2-sphere has a Riemannian metric on M such that there exists a point $p \in M$ satisfying $\widetilde{C}(p) \cap \widetilde{Q}(p) = \varnothing$. On the other hand, Innami, Shiohama and Soga [Innami et al. 2012] have proved that if a complete Riemannian manifold M has a pole p, i.e., $C(p) = \varnothing$, then $\widetilde{C}(q) \cap \widetilde{Q}(q) \neq \varnothing$ for any point $q \in M$ with $C(q) \neq \varnothing$. Ozols [1974] has given a description of $C(p)$ locally around a nonconjugate cut point $q \in C(p)$ as an intersection of a finite number of smooth $(n-1)$-dimensional manifolds and finitely many open sets given by smooth inequalities ($n = \dim M$). Ozols' structure theorem is applicable to all cut points of p in a manifold with Weinstein metric, i.e., $\widetilde{C}(p) \cap \widetilde{Q}(p) = \varnothing$ for the point $p \in M$. Itoh and Sakai [2007] have given the topological structure theorem of a compact manifold with Weinstein metric, using the distance function from p as a Morse function.

The structure of cut loci has been studied in [Warner 1965; 1967; Weinstein 1968; Ozols 1974; Itoh and Tanaka 1998; Itoh and Sakai 2007] and so on. The differentiability and Lipschitz continuity properties of cut loci and the distance functions are studied in [Castelpietra and Rifford 2010; Figalli et al. 2011; Hebda 1994; Itoh 1996; Itoh and Tanaka 2001a; 2001b; Rifford 2004; Tanaka 2003] and so on. As was seen in [Gluck and Singer 1978; Hebda 1983; Itoh and Sabau 2016; Myers 1935; Sabau and Tanaka 2016] and so on, the structure of a cut locus is very complicated.

If $\widetilde{U}_p = \{rv \mid v \in S_pM, 0 \leq r < \rho(v)\}$, then $\exp_p : \widetilde{U}_p \to U_p := M \setminus C(p)$ is a diffeomorphism. Moreover, Ozols [1976] has given a direction preserving diffeomorphism from $\widetilde{U}_p$ onto the open unit ball $B(1)$ in $\mathbb{R}^n$, $n = \dim M$, when M is a compact Riemannian manifold. Then, $M \setminus C(p)$ is diffeomorphic to $B(1)$. Obviously, $\partial \widetilde{U}_p = \widetilde{C}(p)$ where $\partial \widetilde{U}_p$ is the boundary of $\widetilde{U}_p$ in T_pM. Set $\widetilde{U}_p^c = \widetilde{U}_p \cup \widetilde{C}(p)$. We think the *covering* domain $\exp_p : \widetilde{U}_p^c \to M$ to be the development of a Finsler manifold M. When those points in the set ${\exp_p}^{-1}(q)$ are identified for any point $q \in M$, we regard the quotient space $\widetilde{U}_p^c / \exp_p$ as the Finsler manifold M.

We study how to draw $\tilde{c}(s) := {\exp_p}^{-1}(c(s))$ in $\widetilde{U}_p^c$ for a curve $c : [0, 1] \to M$. The problem arises in the case that $c(s_0) \in C(p)$ for some s_0, because ${\exp_p}^{-1}(c(s_0))$ may not be one point. The distribution of ${\exp_p}^{-1}(q)$ in $\widetilde{C}(p)$ for points $q \in C(p)$ is the key to investigate the topological and metrical structure of M.

Ozols' theorem [1974] and Itoh and Sakai's method [2007] show us that, under the condition $\widetilde{C}(p) \cap \widetilde{Q}(p) = \varnothing$, a certain neighborhood of $q \in C(p)$ is decomposed into finitely many sets through each of which there passes a unique minimal geodesic from p to q, and each of those sets are distributed isometrically at corresponding points $\tilde{q}$ contained in the inverse image ${\exp_p}^{-1}(q)$. Here we note that the unit speed minimal geodesics from p to q in M are denoted by $\gamma_v(t) = \exp_p(tv)$, $v = \tilde{q}/F(p, \tilde{q})$, $\tilde{q} \in {\exp_p}^{-1}(q)$. Namely, the Voronoi diagram D

of the negative tangent vectors of all minimal geodesics from p to q in T_qM makes the arrangement of those regions to the points $\tilde{q}$ with $q = \exp_p(\tilde{q})$ in $\widetilde{C}(p) \subset T_pM$. After arrangement, the vectors in those regions are pointing to the inside of $\widetilde{U}_p^c$ at those points $\tilde{q}$.

If $q \in \exp_p(\widetilde{C}(p) \cap \widetilde{Q}(p))$, then these decomposition and arrangement may be more complicated. To describe what happens around q, we use the notion of "limiting tangent line". Let $c'(s)$ be the tangent vector of c at s and $\operatorname{Span}(c'(s))$ the one-dimensional subspace spanned by $c'(s)$ of $T_{c(s)}M$, which is called a *tangent line* of c at s. We say that c has a *limiting tangent line* T at $s = s_0$ if $\operatorname{Span}(c'(s))$ converges to the one-dimensional subspace T of $T_{c(s_0)}M$ as $s \to s_0$.

The north pole p and the south pole q in the unit sphere $S^2(1)$ give a simple example (see Example 7.5): the tangent space at q is decomposed into individual tangent vectors and arranged in each inward normal vector of the circle $S^1(\pi)$ with center 0 and radius π in $T_pS^2(1)$. It is natural to ask how the curves approaching a point $\tilde{q}$ in $S^1(\pi)$ are mapped by $\exp_p$, if their limiting tangent lines are not orthogonal to $S^1(\pi)$ at $\tilde{q}$.

If $q \in \exp_p(\widetilde{C}(p) \cap \widetilde{Q}(p))$ is an end cut point in a surface M, then at q the tangent vector of the minimal geodesic γ_{v_0}, $v_0 \in S_pM$, from p to q is often the limiting tangent line of $C(p)$ at q. Hartman [1964] has stated without proof that $\lambda'(v_0) = 0$ (for the proof, see [Shiohama et al. 2003, Lemma 4.2.3, p. 142]). These facts imply that the half plane bounded by the orthogonal line to v_0 through ${\exp_p}^{-1}(q) =: t_0v_0$ is mapped to the whole tangent space T_qM except for ${\gamma_{v_0}}'(t_0)$ at q by $d\exp_p|_{t_0v_0}$ as the limiting tangent vectors.

Let $\tilde{q} \in \widetilde{C}(p) \cap \widetilde{Q}(p)$ and $q := \exp_p(\tilde{q})$. What happens on the curves approaching q and $\tilde{q}$? Let $\tilde{c}(s)$ and $c(s)$ be curves such that $\tilde{c}(0) = \tilde{q}$, $c(0) = q$ and $\exp_p(\tilde{c}(s)) = c(s)$. In this paper we study how the behaviors of $\tilde{c}$ and c are related. In Section 2, using the first variation formula, we show how to find a converging point $\lim_{s\to 0} {\exp_p}^{-1}(c(s))$ for a curve c with $c(0) \in C(p)$. In Section 3, we study the relation between the tangent vectors of $\tilde{c}(s)$ and $c(s)$ at $s = 0$. Theorem 1.1 is the two-dimensional case of our investigation. We say (see [Gibson 2001, p. 91]) that a function $\varphi(s)$ has a *zero of finite multiplicity m* at s_0 when

$$\varphi(s_0) = 0, \quad \varphi'(s_0) = 0, \quad \dots, \quad \varphi^{(m-1)}(s_0) = 0, \quad \varphi^{(m)}(s_0) \neq 0.$$

Theorem 1.1. *Let M be a geodesically forward complete Finsler surface and $p \in M$. Let $q \in C(p) \cap Q(p)$ be conjugate to p along a minimal geodesic $\gamma_{\theta_0} : [0, d(p,q)] \to M$ from p to q where (t, θ) is a polar coordinate system of T_pM. Assume that $\varphi(\theta) = \lambda(\theta) - \lambda(\theta_0)$ has a zero of finite multiplicity m at $\theta = \theta_0$. Let $c(s)$ be a curve emanating from q such that $c(s) \in M \setminus C(p)$ except for $q = c(0)$ and $\tilde{c}(s)$ the curve satisfying $c(s) = \exp_p(\tilde{c}(s))$. Then the following are true.*

(1) *If the limiting tangent line of* $\tilde{c}(s)$ *at* $s = 0$ *is not tangent to the circle with center* 0 *and radius* $d(p, q)$, *then the limiting tangent line of* $c(s)$ *is tangent to* γ_{θ_0} *at* q.

(2) *If the limiting tangent line of* $c(s)$ *at* $s = 0$ *is not tangent to* γ_{θ_0} *at* q, *then the limiting tangent line of* $\tilde{c}(s)$ *at* $s = 0$ *is tangent to the circle with center* 0 *and radius* $d(p, q)$.

Moreover, if $\tilde{c}(\theta) = (t(\theta), \theta)$ *and* $c'(\theta_0) \notin \mathrm{Span}(\gamma_{\theta_0}{}'(\lambda(\theta_0)))$, *then* $t(\theta) - t(\theta_0)$ *has a zero of multiplicity* $1 + m$ *at* θ_0.

The north pole p and the south pole q in the unit sphere do not satisfy the assumption in Theorem 1.1. However, an end cut point in a surface may satisfy the assumption. We will prove this theorem under a more detailed calculation of the high dimensional case. After we prepare some notations to be used in our discussion, the result will be stated (see Theorem 3.4).

The set of tangent vectors pointing to the interior of $\widetilde{U}_p^c$ at a point $\tilde{q} \in \widetilde{C}(p)$ is considered to be a part of the tangent space $T_{\exp_p(\tilde{q})}M$. In Section 4, for a cut point q, we see how to take $T_q M$ to pieces of convex cones with vertex 0 and find how to arrange those cones at points of $\exp_p{}^{-1}(q)$. After these investigations, for a curve $c(s)$ crossing $C(p)$, we consider what $\tilde{c}(s)$ should be. We propose the notion of *pull back* curves $\tilde{c}(s)$ which satisfy $\exp_p(\tilde{c}(s)) = c(s)$. They are not continuous, in general.

We consider the *pseudo-Finsler metric* F^* on $T_p M$ which is the pullback of F by $\exp_p$, i.e., $F^*(x, y) = F(\exp_p(x), d\exp_p|_x(y))$ for any $y \in T_x T_p M$ and $x \in T_p M$. It follows that $F^*(x, y) = 0$ if and only if $d\exp_p|_x(y) = 0$. There exists a nonzero vector $y \in T_x T_p M$ such that $F^*(x, y) = 0$ if and only if $x \in \widetilde{C}(p) \cap \widetilde{Q}(p)$. Let d^* denote the pseudodistance induced by F^*. We study the relation between d and d^*. In Section 5, we show the relation between M and $\widetilde{U}_p^c/\exp_p$ as distance spaces, using *pull back* curves and d^*.

We say that $C(p)$ is *differentiable* at $q \in C(p)$ if the tangent vector space $T_q C(p)$ of $C(p)$ is defined at q. In Section 6, we study how the set $\exp_p{}^{-1}(q)$ lies in $\widetilde{C}(p)$ when the cut locus $C(p)$ is differentiable at $q \in C(p)$. The Klingenberg lemma and the generalized Berger–Omori theorem proved in [Innami et al. 2019] (see [Berger 1960; 1961; Klingenberg 1959; Nakagawa and Shiohama 1970a; 1970b; Omori 1968] also) suggest us that it is homeomorphic to a sphere (see Theorem 6.5).

In Section 7, we give some examples which help us to understand the discussions and results in this paper.

2. Directional differentiation of a distance function

Let (M, F) be a geodesically forward complete Finsler manifold without boundary. For $y \in T_x M$ let ω_y denote a co-vector in $T_x M^*$ such that $\omega_y(v) = g_y(y, v)$ for any vector $v \in T_x M$ where g_y is the *fundamental tensor* induced by F. Let $f(q) :=$

$d(p, q)$ for all $q \in M$. Then f^2 is C^∞ nearby p and C^1 at p. Actually, $d(q, \cdot)^2$ are C^2 on M for all $q \in M$ if and only if F is Riemannian (see [Shen 2001, Proposition 11.3.3]).

Let $A_p(q)$ be the set of all tangent vectors at q which are the tangent vectors of all constant speed minimal geodesics from p to q and let $A_p(q)^s = A_p(q) \cap S_q M$ where $S_q M$ is the unit sphere with center 0 in $T_q M$. From the first variation formula, the distance function from a point is directionally differentiable. Sabau and Tanaka [2016] have proved this fact of the Finsler manifold version, using second order remainder term. Here we give a slightly modified proof, replacing the second order remainder term by the mean value theorem. Then we use only C^1 differentiability without second order derivatives.

Lemma 2.1 [Sabau and Tanaka 2016]. *Let $c : [0, 1] \to M$ be a curve of class C^1 such that $c(0) = q$ and $c'(0) = w \in T_q M$. We then have*

$$\left.\frac{d(f \circ c)}{dt}\right|_{t=0} = \min_{v \in A_p(q)^s} \omega_v(w).$$

Proof. Let $\gamma : [0, f(q)] \to M$ be a unit speed minimal geodesic from p to q, and let $H : (-\varepsilon, \varepsilon) \times [0, f(q)] \to M$ be a variation of γ through piecewise smooth curves such that $H(0, s) = \gamma(s)$, $H(t, 0) = p$, $H(t, f(q)) = c(t)$. Then $\partial H(0, f(q))/\partial t = w$. If $\gamma_t(s) := H(t, s)$ and $L(\gamma_t)$ is the length of γ_t, then $L(\gamma_t) \geq f \circ c(t)$. Hence, it follows from the first variation formula (see [Shen 2001, equation (5.6)]) that

$$g_{\gamma'(f(q))}(\gamma'(f(q)), w) \geq \limsup_{t \to 0} \frac{f \circ c(t) - f(q)}{t},$$

meaning that

$$\min_{v \in A_p(q)} \omega_v(w) \geq \limsup_{t \to 0} \frac{f \circ c(t) - f(q)}{t}. \tag{1}$$

Assume that t_j is a sequence converging to 0 such that

$$\lim_{j \to \infty} \frac{f \circ c(t_j) - f(q)}{t_j} = \liminf_{t \to 0} \frac{f \circ c(t) - f(q)}{t}$$

and a sequence of minimal geodesics $\gamma_j : [0, f(c(t_j))] \to M$ from p to $c(t_j)$ converges to a minimal geodesic γ from p to q. For a sufficiently small $\delta > 0$, we have

$$f(q) \leq f(\gamma_j(f(c(t_j)) - \delta)) + d(\gamma_j(f(c(t_j)) - \delta), q),$$

and, hence,

$$-d(\gamma_j(f(c(t_j)) - \delta), q) \leq f(\gamma_j(f(c(t_j)) - \delta)) - f(q).$$

Thus we have

$$\begin{aligned}
&d(\gamma_j(f(c(t_j))-\delta), \gamma_j(f(c(t_j)))) - d(\gamma_j(f(c(t_j))-\delta), q) \\
&\qquad \le d(\gamma_j(f(c(t_j))-\delta), \gamma_j(f(c(t_j)))) + f(\gamma_j(f(c(t_j))-\delta)) - f(q) \\
&\qquad = f \circ c(t_j) - f(q).
\end{aligned}$$

Let $\alpha_j : [0, d(q, c(t_j))] \to M$ be the unique minimal geodesic from q to $c(t_j)$ for every j. Since $c(t_j)$ converges to q as $j \to \infty$, we may assume that all $\alpha_j([0, d(q, c(t_j))])$ are contained in a convex ball around $\gamma_j(f(c(t_j)) - \delta)$. If $h(t) := d(\gamma_j(f(c(t_j)) - \delta), \alpha_j(t))$ for $t \in [0, d(q, c(t_j))]$, we then have $h(0) = d(\gamma_j(f(c(t_j)) - \delta), q)$ and $h(d(q, c(t_j))) = d(\gamma_j(f(c(t_j)) - \delta), \gamma_j(f(c(t_j))))$. It follows from the first variation formula that there exists a number $s \in (0, d(q, c(t_j)))$ such that

$$h(d(q, c(t_j))) - h(0) = g_{v_j}(v_j, {\alpha_j}'(s)) d(q, c(t_j)),$$

where v_j denotes the tangent vector of the unit speed minimal geodesic from $\gamma_j(f(c(t_j)) - \delta)$ to $\alpha_j(s)$. Thus we have

$$g_{v_j}(v_j, {\alpha_j}'(s)) d(q, c(t_j)) \le f \circ c(t_j) - f(q).$$

If $c_0(t) = {\exp_q}^{-1}(c(t))$ for sufficiently small $t > 0$, then $c_0(0) = 0$, $d(q, c(t)) = F(q, c_0(t))$ and we have

$$\begin{aligned}
\lim_{t \to +0} \frac{d(q, c(t))}{t} &= \lim_{t \to +0} \frac{F(q, c_0(t) - 0)}{t} \\
&= F\left(q, \lim_{t \to +0} \frac{c_0(t) - 0}{t}\right) \\
&= F(q, {c_0}'(0)) = F(q, c'(0)),
\end{aligned}$$

because $d \exp_q |_0$ is the identity map. Since $\lim_{j \to \infty} {\alpha_j}'(s) = c'(0)/F(q, c'(0))$ and $\lim_{j \to \infty} v_j = \gamma'(f(q))$, we have

$$\lim_{j \to \infty} \frac{g_{v_j}(v_j, {\alpha_j}'(s)) d(q, c(t_j))}{t_j} = g_{\gamma'(f(q))}(\gamma'(f(q)), c'(0)).$$

Therefore we conclude that

$$\min_{v \in A_p(q)^s} \omega_v(w) \le \omega_{\gamma'(f(q))}(w) \le \liminf_{t \to 0} \frac{f \circ c(t) - f(q)}{t}.$$

Combining this inequality with (1), we complete the proof of the lemma. □

Let $X_p(w) := \{y \in S_pM \mid \omega_{\gamma_y'(d(p,q))}(w) = \min_{v \in A_p(q)^s} \omega_v(w)\}$. From Lemma 2.1 we see how to map $c(s)$ into T_pM by $\exp_p{}^{-1}$.

Theorem 2.2. *Let $c : [0, 1] \to M$ be a curve of class C^1 such that $c(0) = q \in C(p)$, $c'(0) = w \in T_qM$ and $c(s) \notin C(p)$ for all $s \in (0, 1]$. Let $\tilde{c}(s)$ be the curve in T_pM such that $\exp_p(\tilde{c}(s)) = c(s)$ for all $s \in (0, 1]$. We then have $\lim_{s\to 0} \tilde{c}(s) \in d(p, q)X_p(w)$.*

3. Curves approaching a conjugate cut point

Let (M, F) be a geodesically forward complete Finsler manifold without boundary. Let $Y(t) := d\exp_p|_{tv}(tw)$ for $v \in S_pM$ and $w \in T_pM$. Then, from Lemma 11.2.2 in [Shen 2001], $Y(t)$, $t \in [0, a)$, is the Jacobi field along $\gamma_v(t) = \exp_p(tv)$ with initial condition $Y(0) = 0$ and $D_{\gamma_v'(0)}Y = w$, where D_v is the covariant derivative at p in the direction v (see (5.33) in [Shen 2001]). From Lemma 6.1.1 in [Shen 2001], it satisfies

$$D_{\gamma_v'} D_{\gamma_v'} Y + R_{\gamma_v'}(Y) = 0$$

where $R_{\gamma_v'}(Y) = R(Y, \gamma_v')\gamma_v'$ is the Riemann curvature which is self-adjoint with respect to the fundamental tensor $g_{\gamma_v'}$ induced by F.

Let $e_1, \dots, e_n$ be an orthonormal basis of T_pM with respect to g_v such that $e_n = v$ and let $E_i(t)$ be the parallel vector field along γ_v with $E_i(0) = e_i$ for each $i = 1, \dots, n$. If $Y(t) = \sum_{j=1}^n y_j(t)E_j(t)$, then $Y(t)$ is identified with the column vector

$$Y(t) = \begin{pmatrix} y_1(t) \\ \vdots \\ y_n(t) \end{pmatrix}$$

Under this notation, the covariant derivative $D_{\gamma_v(t)}Y$ is identified with the differential $Y'(t)$ of its column vector $Y(t)$ with respect to t.

Let $Y_i(t) = d\exp_p|_{tv}(te_i) = \sum_{j=1}^n y_{ji}(t)E_j(t)$. If we set

$$Y_i(t) = \begin{pmatrix} y_{1i}(t) \\ \vdots \\ y_{ni}(t) \end{pmatrix} \quad \text{for } i = 1, \dots, n,$$

then the matrix value function $J(t) = (Y_1(t), \dots, Y_n(t))$ satisfies the differential equation of Jacobi type:

$$J''(t) + R(t)J(t) = 0$$

where $R(t) = \big(g_{\gamma_v'(t)}(R_{\gamma_v'(t)}(E_i(t)), E_j(t))\big)$ is a symmetric $n \times n$ matrix.

Lemma 3.1. *With respect to the orthonormal bases* $\{e_1, \dots, e_n\}$ *for* T_pM *and* $\{E_1(t), \dots, E_{n-1}(t), E_n(t) = \gamma_v{}'(t)\}$ *for* $T_{\gamma_v(t)}M$, *the Jacobi field* $Y(t)$ *along* γ_v *is denoted by*

$$Y(t) = (E_1(t) \cdots E_{n-1}(t)\, \gamma_v{}'(t)) J(v,t) \begin{pmatrix} w_1 \\ \vdots \\ w_{n-1} \\ w_n \end{pmatrix}$$

$$= (E_1(t) \cdots E_{n-1}(t)\, \gamma_v{}'(t)) \begin{pmatrix} J_0(v,t) & 0 \\ 0 & t \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_{n-1} \\ w_n \end{pmatrix},$$

for $w = w_1e_1 + \cdots + w_{n-1}e_{n-1} + w_ne_n$, *where* $J_0(v,t)$ *is the matrix value function satisfying that* $J_0(v,0) = 0$ *and* $J_0{}'(v,0) = I$ (I *is the* $(n-1) \times (n-1)$ *identity matrix*).

Lemma 3.2. *Assume that* $Y_1(t_0) = \cdots = Y_k(t_0) = 0$, $0 \le k \le n-1$, *and that* $\operatorname{rank} J(v,t_0) = n-k$, *i.e.*, $\operatorname{rank} d\exp_p|_{t_0v} = n-k$. *Then the set of tangent vectors* $\{Y_1{}'(t_0), \dots, Y_k{}'(t_0), Y_{k+1}(t_0), \dots, Y_{n-1}(t_0)\}$ *spans the orthogonal complement of* $\gamma_v{}'(t_0)$ *in* $T_{\gamma_v(t_0)}M$. *Furthermore,* $Y_i{}'(t_0)$ *and* $Y_j(t_0)$ *are orthogonal for* $i = 1, \dots, k$ *and* $j = k+1, \dots, n-1$.

Proof. For each $m = 1, \dots, n-1$, if $f(t) = g_{\gamma_v{}'(t)}(\gamma_v{}'(t), Y_m(t)) = 0$ for all $t \ge 0$, then $f(0) = 0$, $f'(0) = 0$ and $f''(t) = 0$ for all $t > 0$, and, hence, $g_{\gamma_v{}'(t)}(\gamma_v{}'(t), Y_m(t)) = 0$ for all $t \ge 0$. Therefore, $Y_m(t_0)$ is contained in the orthogonal compliment of $\gamma_v{}'(t_0)$.

We prove that $\{Y_1{}'(t_0), \dots, Y_k{}'(t_0)\}$ is linearly independent. Suppose for indirect proof that $\sum_{i=1}^k a_iY_i{}'(t_0) = 0$ where at least one of a_i's is not zero. Let $e = \sum_{i=1}^k a_ie_i$ and $Y(t)$ the Jacobi field along γ_v such that $Y(0) = 0$ and $Y'(0) = e$. Obviously, Y is not identically zero. However, since a Jacobi field Y satisfies $Y(t_0) = 0$ and $Y'(t_0) = 0$, we have $Y(t) = 0$ identically, a contradiction.

From this and

$$\operatorname{Span}(Y_1(t), \dots, Y_k(t)) = \operatorname{Span}\left(\frac{Y_1(t)}{t-t_0}, \dots, \frac{Y_k(t)}{t-t_0}\right), \quad 0 < t < t_0,$$

we see that $\operatorname{Span}(Y_1(t), \dots, Y_k(t))$ converges to $\operatorname{Span}(Y_1{}'(t_0), \dots, Y_k{}'(t_0))$ as $t \to t_0$.

Since $\operatorname{rank} J(v,t_0) = n-k$, we see that $\{Y_{k+1}(t_0), \dots, Y_n(t_0)\}$ is linearly independent.

We next prove that $Y_i{}'(t_0)$ and $Y_j(t_0)$ are orthogonal for $i = 1, \dots, k$ and $j = k+1, \dots, n-1$. Since both Y_i and Y_j are Jacobi fields along γ_v with

$Y_i(0) = Y_j(0) = 0$, we see that $g_{\gamma_v'(t)}(Y_i'(t), Y_j(t)) - g_{\gamma_v'(t)}(Y_i(t), Y_j'(t))$ is constant for t and zero at $t = 0$. From this, we have

$$g_{\gamma_v'(t_0)}(Y_i'(t_0), Y_j(t_0)) = g_{\gamma_v'(t_0)}(Y_i(t_0), Y_j'(t_0)) = 0$$

because $Y_i(t_0) = 0$. This proves Lemma 3.2. □

Remark 3.3. As a simple application of Lemma 3.2, we get the following well-known fact: If $\gamma_v(t_0)$ is conjugate to $\gamma_v(0)$ along a geodesic γ_v, then there exists a $\delta > 0$ such that no point $\gamma_v(t)$ with $0 < |t - t_0| < \delta$ is conjugate to $\gamma_v(0)$. Because we have $\det(Y_1'(t_0) \cdots Y_k'(t_0)\, Y_{k+1}(t_0) \cdots Y_{n-1}(t_0)) \neq 0$ and

$$\det J(t) = (t - t_0)^k \det(Y_1'(t_1) \cdots Y_k'(t_k)\, Y_{k+1}(t) \cdots Y_{n-1}(t))$$

where, from the mean value theorem,

$$Y_i'(t_i) = \begin{pmatrix} y_{1i}'(t_{1i}) \\ \vdots \\ y_{ni}'(t_{ni}) \end{pmatrix} \quad \text{for } i = 1, \dots, k$$

for some t_{ji} with $|t_{ji} - t_0| < |t - t_0|$, $j = 1, \dots n$.

We investigate how the preimage $\exp_p{}^{-1}(c(s))$ behaves for a curve $c(s)$ if $q = c(0)$ is a conjugate cut point along a geodesic γ_v and $c(s) \in M \setminus C(p)$ for $s \neq 0$.

Let $v_0 = e_n \in S_pM$ and $\{e_1, \dots, e_{n-1}\}$ an orthonormal basis of the orthogonal complement $v_0{}^\perp$ of v_0 with respect to g_{v_0} in T_pM. We use this basis to have a coordinate system for T_pM. Let $(V, \tau^{-1} = (v_1, \dots, v_{n-1}))$ be a local coordinate system of S_pM around v_0, where $\tau : W = \tau^{-1}(V) \subset \mathbb{R}^{n-1} \to V \subset S_pM$, such that $\tau(0) = v_0$ and $d\tau_0$ is an isometry from $T_0\mathbb{R}^{n-1}$ to $T_{v_0}S_pM$ with respect to $g_{v_0}|_{T_{v_0}S_pM}$ and let $(\mathbb{R}_+V, \psi^{-1} = (v_1, \dots, v_{n-1}, t))$ be a polar coordinate system of T_pM such that $\psi(v, t) = t\tau(v)$ for $t > 0$ and $v \in W = \tau^{-1}(V)$. Then we have

$$d\psi_{(0,t)} = \begin{pmatrix} t\,d\tau_0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Assume that $q = \gamma_{\tau(0)}(t_0)$ for $\tau(0) = v_0$. We make a local coordinate system $(U, x_1, \dots, x_n)$ in a tubular neighborhood U around $\gamma_{\tau(0)}$ at q as follows:

(1) $x_1(q) = x_2(q) = \cdots = x_n(q) = 0$.

(2) $x_1(\exp_p(\psi(0, t))) = 0, \dots, x_{n-1}(\exp_p(\psi(0, t))) = 0$ and $x_n(\exp_p(\psi(0, t))) = t_0 - t$.

(3) If $z = \gamma_w(s)$ for $w = \sum_{i=1}^{n-1} a_i E_i(t)$ where γ_w is the geodesic with $\gamma_w'(0) = w$ and $\sum_{i=1}^{n-1} a_i{}^2 = 1$, then

$$(x_1(z), \dots, x_{n-1}(z), x_n(z)) = (sa_1, \dots, sa_{n-1}, t_0 - t).$$

Since $c(s) \in M \setminus C(p)$ for $s \neq 0$, we have $\tilde{c}(s) = \psi^{-1}(\exp_p{}^{-1}(c(s))) = (v(s), t(s))$ such that $t(0) = t_0$ and $v(0) = 0$, i.e., $\tau(v(0)) = v_0$. We study how $t'(s)$ and $v'(s)$ behave as $s \to 0+0$. In the coordinate systems defined as above, let $\exp_p(t\tau(v)) = \exp_p(\psi(v,t)) = (x_1(v,t), \ldots, x_n(v,t))$. If the partial derivative of x_i with respect to v_j is written by $x_{i,j}$, then $d\exp_p|_{t\tau(v)} \circ d\psi_{(v,t)}$ is expressed by

$$d\exp_p|_{t\tau(v)} \circ d\psi_{(v,t)} = \begin{pmatrix} x_{1,1}(v,t) & \cdots & x_{1,n}(v,t) \\ \vdots & & \vdots \\ x_{n,1}(v,t) & \cdots & x_{n,n}(v,t) \end{pmatrix}$$

with respect to the bases $\left\{\frac{\partial}{\partial v_j}, \ldots, \frac{\partial}{\partial v_{n-1}}, \frac{\partial}{\partial t}\right\}$ and $\left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right\}$. Let the column vectors

$$\begin{pmatrix} y_{1i}(v,t) \\ \vdots \\ y_{ni}(v,t) \end{pmatrix}, \quad i = 1, \ldots, n-1,$$

denote the Jacobi fields

$$Y_i(v,t) = \frac{\partial \exp_p(t\tau(v))}{\partial v_i}$$

along $\gamma_{\tau(v)}$ with respect to the basis $\left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right\}$. Then, we have

$$d\exp_p|_{t\tau(v)} \circ d\psi_{(v,t)} = \begin{pmatrix} y_{11}(v,t) & \cdots & y_{1n-1}(v,t) & x_{1,n}(v,t) \\ \vdots & & \vdots & \vdots \\ y_{n1}(v,t) & \cdots & y_{nn-1}(v,t) & x_{n,n}(v,t) \end{pmatrix}.$$

Here the n-th column vector is the coordinate of $\gamma_{\tau(v)}{}'(t)$. When we assume that there exists a positive integer $k := \nu(v) = \dim\ker d\exp_p|_{\lambda(v)\tau(v)} > 0$ for all v in a neighborhood U of 0, it follows from the implicit function theorem that $\lambda(v) := \lambda(\tau(v))$ is smooth in U. Furthermore, we can choose a coordinate system around q such that

$$\begin{aligned} &d\exp_p|_{\lambda(v)\tau(v)} \circ d\psi_{(v,\lambda(v))} \\ &\quad = \begin{pmatrix} 0 & \cdots & 0 & y_{1k+1}(v,\lambda(v)) & \cdots & y_{1n-1}(v,\lambda(v)) & x_{1,n}(v,\lambda(v)) \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & y_{nk+1}(v,\lambda(v)) & \cdots & y_{nn-1}(v,\lambda(v)) & x_{n,n}(v,\lambda(v)) \end{pmatrix}. \end{aligned}$$

From the mean value theorem, there exists a number $t_{ij}(v,t)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n-1$ between $\lambda(v)$ and t such that

$$x_{i,j}(v,t) = y_{ij}(v,\lambda(v)) + (t - \lambda(v))y_{ij}{}'(v, t_{ij}(v,t)).$$

It should be noted that the covariant derivatives $D_{\gamma_{\tau(v)}'(t)}Y_i$ do not equal the differential $Y_i'(v,t)$ with respect to t, in general. However, because $Y_i(v,\lambda(v))=0$, we have $D_{\gamma_{\tau(v)}'(\lambda(v))}Y_i = Y_i'(v,\lambda(v))$.

We assume that $e_i = d\tau_0\big(\frac{\partial}{\partial v_i}\big)$ for $i=1,\dots,n-1$. Let $w_1 \in \mathrm{Span}(e_1,\dots,e_k)$ and $w_2 \in \mathrm{Span}(e_{k+1},\dots,e_{n-1})$ and $w_3 \in \mathbb{R}$. We briefly write $w=(w_1,w_2)=(w_1,\dots,w_{n-1})$, $w_1=(w_{11},w_{21},\dots,w_{k1})$ and $w_2=(w_{k+12},w_{k+22},\dots,w_{n-12})$.

It follows from Taylor's theorem with integral form of the remainder (see [Warner 1971, Lemma on p. 13]) that

$$\begin{aligned}\lambda(w_1,w_2)-\lambda(0,0)\\ &=\sum_{i_1=1}^{n-1} w_{i_1}\lambda_{i_1}(0,0)+\cdots+\frac{1}{\ell!}\sum_{i_1,\dots,i_\ell=1}^{n-1} w_{i_1}\cdots w_{i_\ell}\lambda_{i_1\cdots i_\ell}(0,0)\\ &\qquad+\frac{1}{\ell!}\sum_{i_1,\dots,i_{\ell+1}=1}^{n-1} w_{i_1}\cdots w_{i_{\ell+1}}\int_0^1(1-s)^\ell\lambda_{i_1\cdots i_{\ell+1}}(sw)\,ds.\end{aligned}$$

Define an integer $m(w)>0$ by

$$m(w)=\min\Big\{\ell>0\;\Big|\;\sum_{i_1,\dots,i_\ell} w_{i_1}\cdots w_{i_\ell}\lambda_{i_1\cdots i_\ell}(0,0)\neq 0,\quad i_j\in\{1,\dots,k\},\, j=1,\dots,\ell\Big\}.$$

We then define a function g by, if $m=m(w)\neq\infty$,

$$g(w_1)=\frac{1}{m!}\sum_{i_1,\dots,i_m} w_{i_1 1}\cdots w_{i_m 1}\lambda_{i_1\cdots i_m}(0,0),$$

where $i_j\in\{1,\dots,k\}$, $j=1,\dots,m$, and $g(w_1)=0$ if $m(w)=\infty$. Then g is a homogeneous function with degree m. Further we define a function f by

$$f(w_1,w_2,s,u)=\lambda((sw_1,uw_2))-\lambda(0,0)-s^m g(w_1)$$

for any $(w_1,w_2)\in T_0\mathbb{R}^{n-1}=\mathbb{R}^k\times\mathbb{R}^{n-k-1}$. From the definition of f, each term contains the parameter u as a factor or the order of s is greater than m. In particular,

$$\lim_{s\to 0}\frac{f(w_1,w_2,s,s^{1+m})}{s^m}=0,$$

or, in other words,

$$\lim_{s\to 0}\frac{\lambda((sw_1,s^{1+m}w_2))-\lambda(0,0)}{s^m}=g(w_1).$$

Hence we have

$$\begin{aligned}
&x_{i,j}(sw_1, uw_2, t)\\
&\quad= y_{ij}(sw_1, uw_2, \lambda(sw_1, uw_2))\\
&\qquad+ (t - \lambda(sw_1, uw_2))y_{ij}{}'(sw_1, uw_2, t_{ij}(sw_1, uw_2, t))\\
&\quad= y_{ij}(sw_1, uw_2, \lambda(sw_1, uw_2))\\
&\qquad+ (t - \lambda(0, 0) - s^m g(w_1) - f(w_1, w_2, s, u))y_{ij}{}'(sw_1, uw_2, t_{ij}(sw_1, uw_2, t)).
\end{aligned}$$

Recall for the next step that $\lambda(0, 0) = t_0$ and $y_{ij}(v, \lambda(v)) = 0$ for $i = 1, \dots, n$ and $j = 1, \dots, k$ where v are near 0.

We first consider a curve

$$\tilde{c}(s) = \psi((sw_1, sw_2, t_0 + sw_3))$$

and set $c(s) = \exp_p(\tilde{c}(s))$. Since $\tilde{c}(0) = \psi((0, 0, t_0))$ and $\tilde{c}'(0) = d\psi((w_1, w_2, w_3))$, we have

$$c'(0) = \sum_{j=k+1}^{n-1} w_{j2} Y_j(0, 0, t_0) - w_3 E_n(t_0).$$

Therefore, we see that

$$c'(0) \in \operatorname{Span}\big(Y_{k+1}(0, 0, t_0), \dots, Y_{n-1}(0, 0, t_0), E_n(0, 0, t_0)\big).$$

Next, we consider a curve $\tilde{c}(s) = \psi((sw_1, s^{1+m}w_2, t_0 + s^{1+m}w_3))$ for $w = (w_1, w_2)$ with $m = m(w) > 0$ and set $c(s) = \exp_p(\tilde{c}(s))$. Then, we have $\tilde{c}(0) = \psi((0, 0, t_0))$ and

$$\tilde{c}'(s) = d\psi((w_1, (1 + m)s^m w_2, (1 + m)s^m w_3)).$$

Moreover, we have

$$\begin{aligned}
&\lim_{s\to 0} \frac{c'(s)}{s^m}\\
&\quad= -g(w_1)\sum_{j=1}^{k} w_{j1} Y_j{}'(0, 0, t_0) + (1 + m)\sum_{j=k+1}^{n-1} w_{j2} Y_j(0, 0, t_0) - (1 + m)w_3 E_n(t_0).
\end{aligned}$$

Since $\{Y_1{}'(0, 0, t_0), \dots, Y_k{}'(0, 0, t_0), Y_{k+1}(0, 0, t_0), \dots, Y_{n-1}(0, 0, t_0)\}$ spans the orthogonal complement of $\gamma_{\tau(0)}{}'(t_0)$ in $T_{\gamma_{\tau(0)}(t_0)}M$, the above vector may become any tangent vector in $T_{\gamma_{v_0}(t_0)}M$. After changing the parameter, we have a curve

$$\tilde{c}(s) = \psi\big(((1 + m)s)^{1/(1+m)} w_1, (1 + m)sw_2, t_0 + (1 + m)sw_3\big).$$

The image of this curve is the same as the previous one. We note that $\tilde{c}(s)$ is not differentiable at $s = 0$, but $c(s)$ is differentiable at $s = 0$.

Summarizing the discussion so far, we get the following theorem.

Theorem 3.4. *Let M be a geodesically forward complete Finsler manifold without boundary and $p \in M$. Let $\gamma(t) = \exp_p(tv)$ be a minimal geodesic from p to q such that $q = \gamma(t_0)$ is conjugate to p along γ. Suppose $k = \dim \ker d \exp_p |_{t_0 v} \geq 1$ is constant around v in $S_p M$. Let Jacobi vector fields $Y_1, \dots, Y_{n-1}$ be defined as above such that $Y_1(t_0) = \cdots = Y_k(t_0) = 0$. Then we have an orthogonal sum $T_q M = W_1 + W_2 + \mathrm{Span}(\gamma'(t_0))$, where $W_1 = \mathrm{Span}({Y_1}'(t_0), \dots, {Y_k}'(t_0))$ and $W_2 = \mathrm{Span}(Y_{k+1}(t_0), \dots, Y_{n-1}(t_0))$. Let $c(s)$ and $\tilde{c}(s)$, $s \in (0, 1)$, be smooth curves such that $c(s) = \exp_p(\tilde{c}(s)) \in M \setminus C(p)$ for all $s \in (0, 1)$, and $\lim_{s \to 0} c(s) = q$ and $\lim_{s \to 0} \tilde{c}(s) = t_0 v$. Then the following are true.*

(1) *If $\tilde{c}(s)$ is differentiable at $s = 0$, then $c'(0) \in W_2 + \mathrm{Span}(\gamma'(t_0))$.*

(2) *Suppose that $c'(0)$ exists and $w \neq 0$ is the projection of $c'(0)$ to W_1. If there exists a vector $w_1 \in \mathbb{R}^k$ such that $w = -g(w_1) \sum_{j=1}^{k} w_{j1} {Y_j}'(t_0)$, then $\lim_{s \to 0} \mathrm{Span}(\tilde{c}'(s)) \in \ker d \exp_p |_{t_0 v}$.*

4. Sending curves into the tangent space by ${\exp_p}^{-1}$

Let (M, F) be a geodesically forward complete Finsler manifold without boundary. For $q \in C(p)$ let $T_q C(p)$ be the *tangent cone* of $C(p)$ at q, i.e., $v \in T_q C(p) \setminus \{0\}$ if and only if there exists a sequence of vectors $v_j \in T_q M$ converging to 0 with $\exp_q(v_j) \in C(p)$ such that $v/F(q, v) = \lim_{j \to \infty} v_j / F(q, v_j)$. We do not know whether $\{-v \mid v \in A_p(q)\} \cap T_q C(p) = \varnothing$ is true or not.

The *Voronoi region* $V(v)$ for $v \in A_p(q)^s$ in $T_q M$ is defined by

$$V(v) = \{w \in T_q M \mid \omega_v(w) < \omega_u(w) \text{ for all } u \in A_p(q)^s \text{ with } u \neq v\}.$$

Suppose $A_p(q)^s$ consists of more than one vector. If

$$H_v(u) := \{w \in T_q M \mid \omega_v(w) - \omega_u(w) < 0\}$$

for any $u \in A_p(q)^s$ with $u \neq v$, then $V(v) = \bigcap_{u \in A_p(q)^s, u \neq v} H_v(u)$. If $A_p(q)^s = \{v\}$, we then set $V(v) = T_q M \setminus \{0\}$. It follows that $w \in V(v)$ if and only if there exists the unique vector $v \in A_p(q)^s$ such that $\omega_v(w) = \min\{\omega_u(w) \mid u \in A_p(q)^s\}$.

Lemma 4.1. *The following are true.*

(1) *$-v \in V(v)$ for any $v \in A_p(q)^s$. If $v \in A_p(q)^s$ and $A_p(q)^s \neq \{v\}$, then $v \notin V(v)$.*

(2) *$V(v)$ is a cone for any $v \in A_p(q)^s$, i.e., $\mu w \in V(v)$ for any $w \in V(v)$ and any $\mu > 0$.*

(3) *$V(v) \cup \{0\}$ is convex, i.e., $\mu w_1 + (1 - \mu) w_2 \in V(v) \cup \{0\}$ for any $\mu \in [0, 1]$ if $w_1, w_2 \in V(v)$.*

(4) *$V(v) \cap V(u) = \varnothing$ and $\overline{V}(v) \cap \overline{V}(u) \subset \ker(\omega_v - \omega_u)$ for any $u, v \in A_p(q)^s$ with $u \neq v$.*

(5) $\bigcup_{v \in A_p(q)^s} V(v)$ *is dense in* T_qM. *In particular*, $\bigcup_{v \in A_p(q)^s} \overline{V(v)} = T_qM$.

Proof. From the Cauchy–Schwarz inequality (see [Shen 2001, Lemma 1.2.6]: $\omega_u(w) \le F(u)F(w)$ for all $w \in T_qM$ with equality holding if and only if $w = \mu u$ for some $\mu \ge 0$), we have

$$\omega_u(-v) \ge -F(v)F(u) \ge -1 = \omega_v(-v)$$

for all $u \in A_p(q)^s$. Suppose the equality $\omega_u(-v) = \omega_v(-v)$ holds. Then we have $\omega_u(v) = \omega_v(v) = F(v)F(u)$. Hence $v = \mu u$ for some $\mu \ge 0$. Then, $1 = \omega_u(v) = g_u(u, \mu u) = \mu$, meaning $u = v$. Therefore, the equality does not hold if $u \ne v$.

If there exists a vector $u \in A_p(q)^s$ with $u \ne v$, then $\omega_v(v) = 1 = F(v)F(u) > \omega_u(v)$, meaning that $v \notin V(v)$. These prove (1).

Since ω_v is the homogeneous with degree 1 for every $v \in A_p(q)^s$, we have (2).

We prove (3). If $A_p(q)^s = \{v\}$, we then have nothing to prove because $V(v) = T_qM \setminus \{0\}$. Assume that $v \in A_p(q)^s$ and $A_p(q)^s \ne \{v\}$. Let $w_1, w_2 \in V(v)$. We then have $\omega_v(w_1) < \omega_u(w_1)$ and $\omega_v(w_2) < \omega_u(w_2)$ for any $u \in A_p(q)^s$ with $u \ne v$. We may assume that $\mu w_1 + (1-\mu)w_2 \ne 0$ for $\mu \in [0, 1]$. Then, we have

$$\begin{aligned}\omega_v(\mu w_1 + (1-\mu)w_2) &= \mu\omega_v(w_1) + (1-\mu)\omega_v(w_2)\\ &< \mu\omega_u(w_1) + (1-\mu)\omega_u(w_2)\\ &= \omega_u(\mu w_1 + (1-\mu)w_2)\end{aligned}$$

because of (2), proving that $\mu w_1 + (1-\mu)w_2 \in V(v)$.

From the definition of $V(v)$, property (4) is a direct consequence.

We prove (5). Suppose that there exists $w \in T_qM \setminus \bigcup_{v \in A_p(q)^s} V(v)$. It follows from (1) that $w \ne -v$ for any $v \in A_p(q)^s$. We assume that $\omega_v(w) = \inf\{\omega_u(w) \mid u \in A_p(q)^s\}$ for some $v \in A_p(q)^s$. Set $w(\varepsilon) = w - \varepsilon v$ for any $\varepsilon > 0$. Then, for any $u \in A_p(q)^s$ with $u \ne v$, we have

$$\begin{aligned}\omega_u(w(\varepsilon)) &= \omega_u(w) + \varepsilon\omega_u(-v)\\ &> \omega_v(w) - \varepsilon\\ &= \omega_v(w) + \varepsilon\omega_v(-v)\\ &= \omega_v(w(\varepsilon)).\end{aligned}$$

From this, we see that $w(\varepsilon) \in V(v)$, and, hence, $w \in \overline{V(v)} \subset \overline{\bigcup_{v \in A_p(q)^s} V(v)}$. This implies that $\overline{\bigcup_{v \in A_p(q)^s} V(v)} = T_qM$. Thus, $\bigcup_{v \in A_p(q)^s} \overline{V(v)} = T_qM$. □

Let $W_q := \bigcup_{v \in A_p(q)^s} V(v)$ which is dense in T_qM at $q \in C(p)$.

Remark 4.2. We do not know whether $W_q \cap T_qC(p) = \varnothing$ is true or not.

Let $B_f(q, \varepsilon)$ be the forward distance ball with center q and radius ε.

Lemma 4.3. *Let $q \in C(p)$. If q is not a conjugate cut point, then $W_q \cap T_q C(p) = \varnothing$. In particular, for any smooth curve $c : [0, 1] \to M$ with $c(0) = q$ and $c'(0) \in W_q$, there exists a number $\delta > 0$ such that $c(s) \in M \setminus C(p)$ for all $s \in (0, \delta)$.*

Proof. Since q is not a conjugate cut point, there exists a number $\varepsilon > 0$ such that $B_f(q, \varepsilon) \cap C(p)$ is a union of smooth hypersurfaces with boundaries (see [Ozols 1974]). This implies that $w \in W_q$ if and only if $w \notin T_q C(p)$. □

Lemma 4.4. *If $w \in W_q \setminus T_q C(p)$, then there exist the unique point $x \in \widetilde{U}_p^c$ and a curve $\tilde{c} : [0, \varepsilon) \to T_p M$ such that $\tilde{c}(0) = x$, $\tilde{c}$ is of class C^∞ on $(0, \varepsilon)$ and*

$$\left.\frac{d \exp_p \circ \tilde{c}}{ds}\right|_{s=0} = w.$$

Proof. Assume that $w \in V(v)$ for $v \in A_p(q)^s$. Let $\gamma : [0, d(p, q)] \to M$ be the minimal geodesic from p to q such that $\gamma'(d(p, q)) = v$ and let $c : [0, \varepsilon) \to M$ be a curve such that $c'(0) = w$. From Lemma 2.1, it follows that

$$\left.\frac{dd(p, c(s))}{ds}\right|_{s=0} = \min\{\omega_u(w) \mid u \in A_p(q)^s\} = \omega_v(w).$$

This implies that a sequence of minimal geodesics $T(p, c(s))$ from p to $c(s)$ converges to γ as $s \to 0+0$. Then $x = d(p, q)\gamma'(0)$. Since $w \notin T_q C(p)$, there exists a unique minimal geodesic $T(p, c(s))$ from p to $c(s)$ for a sufficiently small $s > 0$. If $\tilde{c}(s) = d(p, c(s))y(s)$, where $y(s)$ are the initial tangent vectors of $T(p, c(s))$ at p, then $c(s) = \exp_p(\tilde{c}(s))$ for $s \in [0, \varepsilon)$. □

Remark 4.5. As was seen in Theorem 3.4, $\tilde{c}'(0)$ may not exist: let M be a surface. Then we know from [Shiohama and Tanaka 1996] that $C(p)$ is locally a tree. Suppose there exists an end point q of $C(p)$ such that the sufficiently short edge e ending at q is smooth. Then q is a point conjugate to p along a minimal geodesic γ_v from p to q. If γ_v is the unique minimal geodesic, then the edge e has two lifts $\tilde{e}_1$ and $\tilde{e}_2$ in $T_p M$ by $\exp_p$ such that $\tilde{e}_1 \cap \tilde{e}_2 = \{\lambda(v)v\}$ where $\exp_p(\lambda(v)v) = q$. They are tangent at $\lambda(v)v$, i.e., they are linearly dependent because $\lambda'(v) = 0$ as mentioned in Section 1. Let N be the tangent space of the circle with center origin and radius $d(p, q)$ in $T_p M$. Then $\ker d \exp_p|_x = N$. Therefore, we can not find any $\tilde{w}$ such that $d \exp_p|_x(\tilde{w}) = w$ if w is a tangent vector at q orthogonal to ${\gamma_v}'(d(p, q))$.

Theorem 4.6. *Let $c(s)$, $s \in [a, b]$, be a curve of class C^1 in M such that $c'(a) \neq 0$ and $c(s) \notin C(p)$ for all $s \in (a, b]$. Then there exists the unique curve $\tilde{c}(s)$, $s \in [a, b]$, in $\widetilde{U}_p^c$ such that $\exp_p(\tilde{c}(s)) = c(s)$ for all $s \in [a, b]$.*

Proof. Since $\exp_p : \widetilde{U}_p \to U_p$ is a diffeomorphism, $\tilde{c}(s) = (\exp_p|_{\widetilde{U}_p})^{-1}(c(s))$ satisfies the required condition in $s \in (a, b]$. When $c(a) \notin C(p)$, $\tilde{c}(a) = (\exp_p|_{\widetilde{U}_p})^{-1}(c(a))$ is also defined. Therefore, $\tilde{c}(s)$, $s \in [a, b]$, is the unique curve mentioned in the theorem.

Assume that $q := c(a) \in C(p)$. Let $\gamma_s : [0, d(p, c(s))] \to M$ be minimal geodesics from p to $c(s)$ for all $s \in (a, b]$. Assume that $v \in A_p(q)^s$ is an accumulation tangent vector of $\gamma_s{}'(d(p, c(s)))$ as $s \to 0$. If $w := c'(a) \in V(v) \setminus T_q C(p)$, then we can have $x \in T_p M$ and a curve $\tilde{c}$ mentioned in this theorem as was seen in Lemma 4.4 and $\gamma_s{}'(d(p, c(s)))$ converges to v as $s \to a + 0$.

Assume that $w \in \overline{V}(v) \setminus V(v)$. For any sufficiently small $\varepsilon > 0$, let $c_\varepsilon(s) = \gamma_s(d(p, c(s)) - \varepsilon s)$ for $s \in (a, b]$. Then we have $c_\varepsilon(s) \notin C(p)$ and $c_\varepsilon{}'(a) = w - \varepsilon v$. Since $c_\varepsilon{}'(a) \in V(v) \setminus T_q C(p)$, we have a curve $\tilde{c}_\varepsilon(s)$ such that $\exp_p(\tilde{c}_\varepsilon(s)) = c_\varepsilon(s)$. Thus, $\exp_p(\tilde{c}(s)) = c(s)$ as $\varepsilon \to 0 + 0$. □

We call a map $\tilde{c} : [a, b] \to \widetilde{U}_p^c$ a *pull back curve* through $\exp_p$ (briefly, *pb-curve*) if there exists a set of numbers $\{a_\lambda \mid \lambda \in \Lambda\} \cup \{b_\lambda \mid \lambda \in \Lambda\}$ in $[a, b]$ satisfying the following conditions:

(1) $(a_\lambda, b_\lambda) \cap (a_{\lambda'}, b_{\lambda'}) = \varnothing$ for $\lambda, \lambda' \in \Lambda$ with $\lambda \neq \lambda'$ and $\tilde{c}^{-1}(\widetilde{U}_p^c \setminus \widetilde{C}(p)) \setminus \{a, b\} = \bigcup_{\lambda \in \Lambda} (a_\lambda, b_\lambda)$. In particular, we have $\tilde{c}\big((a, b) \setminus \bigcup_{\lambda \in \Lambda} (a_\lambda, b_\lambda)\big) \subset \widetilde{C}(p)$.

(2) Let $\tilde{c}_\lambda := \tilde{c}|_{[a_\lambda + 0, b_\lambda - 0]}$ for every $\lambda \in \Lambda$. Then $\tilde{c}_\lambda$ is continuous on $[a_\lambda, b_\lambda]$ and piecewise smooth on (a_λ, b_λ) for each $\lambda \in \Lambda$.

(3) Let $c := \exp_p \circ \tilde{c}$. Then $c : [a, b] \to M$ is a piecewise smooth curve.

Since c is piecewise smooth, condition (3) implies that $c(b_\lambda - 0) = c(a_{\lambda'} + 0)$ if $b_\lambda = a_{\lambda'}$. Since the pb-curve $\tilde{c}$ is not assumed to be continuous, it may happen that $\tilde{c}(b_\lambda - 0) \neq \tilde{c}(a_{\lambda'} + 0)$ even if $c(b_\lambda - 0) = c(a_{\lambda'} + 0) \in C(p)$.

We say that a nonconjugate cut point q of p is *normal* if exactly two minimal geodesics from p to q exist in M. It follows from the implicit function theorem that the set of all normal cut points of p makes a smooth hypersurface of M. Furthermore, Itoh and Tanaka [1998] proved that the Hausdorff dimensions of the sets of all conjugate cut points of p and all nonnormal cut points of p are not greater than $\dim M - 2$.

Lemma 4.7. *Let $c : [a, b] \to M$ be a piecewise smooth curve. Assume that c does not pass through any conjugate cut point and any nonnormal cut point of p. Then there exists a pb-curve $\tilde{c} : [a, b] \to \widetilde{U}_p^c$ such that* $\exp_p(\tilde{c}(t)) = c(t)$ *and $\tilde{c}$ is a union of piecewise smooth curves.*

Proof. Since $M \setminus C(p)$ is an open set, there exists a set of numbers

$$\{a_\lambda \mid \lambda \in \Lambda\} \cup \{b_\lambda \mid \lambda \in \Lambda\}$$

such that

$$c^{-1}(M \setminus C(p)) \setminus \{a, b\} = \bigcup_{\lambda \in \Lambda} (a_\lambda, b_\lambda).$$

Then we have curves $\tilde{c}_\lambda : (a_\lambda, b_\lambda) \to \widetilde{U}_p^c \setminus \widetilde{C}(p)$ such that $\exp_p \circ \tilde{c}_\lambda = c|_{(a_\lambda, b_\lambda)}$ for all $\lambda \in \Lambda$. The domains of these curves are extended to their closures because of Lemma 4.3 and Theorem 4.6.

Each connected component ℓ of $c([a, b]) \cap C(p)$ has two curves $\tilde{g}$ and $\tilde{h}$ in $\widetilde{C}(p)$, i.e., $\ell = \exp_p \circ \tilde{g} = \exp_p \circ \tilde{h}$, since ℓ is a piecewise smooth curve in a smooth hypersurface in M consisting of normal cut points of p. We choose one of $\tilde{g}$ and $\tilde{h}$ for a pb-curve of ℓ. Thus the union of those curves and $\tilde{c}_\lambda$, $\lambda \in \Lambda$, makes a pb-curve $\tilde{c}$ of c. □

5. The relation between the distances induced by F and F^*

Let (M, F) be a geodesically forward complete Finsler manifold without boundary. For a piecewise smooth curve $\tilde{c}(t)$, $t \in [a, b]$, in $\widetilde{U}_p^c$ the length $L(\tilde{c})$ of $\tilde{c}$ is defined by

$$L(\tilde{c}) = \int_a^b F^*(\tilde{c}(t), \tilde{c}'(t))\, dt.$$

It follows that $L(\tilde{c}) = L(\exp_p \circ \tilde{c})$. Let $\Omega(x, y)$ denote the set of all piecewise smooth curves in $\widetilde{U}_p^c$ from x to y for $x, y \in \widetilde{U}_p^c$ and $\Omega_0(x, y)$ the set of all pb-curves in $\widetilde{U}_p^c$ whose image by $\exp_p$ connects $\exp_p(x)$ to $\exp_p(y)$ in M. Obviously, we have $\Omega(x, y) \subset \Omega_0(x, y)$. For $\tilde{c} \in \Omega_0(x, y)$, we have

$$L(c) = L(\tilde{c}) = \sum_{\lambda \in \Lambda} \int_{a_\lambda}^{b_\lambda} F^*(\tilde{c}_\lambda(t), \tilde{c}_\lambda{}'(t))\, dt + \int_{[a,b] \setminus \cup_{\lambda \in \Lambda}(a_\lambda, b_\lambda)} F^*(\tilde{c}(t), \tilde{c}'(t))\, dt,$$

from the definition of a pb-curve.

We define the pseudodistances $d^*(x, y)$ and $d_0^*(x, y)$ from x to y by

$$d^*(x, y) = \inf\{L(\tilde{c}) \mid \tilde{c} \in \Omega(x, y)\}$$
$$d_0^*(x, y) = \inf\{L(\tilde{c}) \mid \tilde{c} \in \Omega_0(x, y)\}.$$

It follows that $d^*(x, y) \geq d_0^*(x, y) \geq d(\exp_p(x), \exp_p(y))$ for any $x, y \in \widetilde{U}_p^c$. It may happen that $d^*(x, y) = 0$ for $x \neq y$ when there exists a curve $c(t)$, $t \in [a, b]$, from x to y in $\widetilde{C}(p)$ such that $d\exp_p(\tilde{c}'(t)) = 0$ for $t \in [a, b]$.

Lemma 5.1. *For $d_0^*(x, y)$ as above, we have $d_0^*(x, y) = d(\exp_p(x), \exp_p(y))$ for any $x, y \in \widetilde{U}_p^c$.*

Proof. We prove that

$$d_0^*(x, y) \leq d(\exp_p(x), \exp_p(y)).$$

For any $\varepsilon > 0$ let $c : [a, b] \to M$ be a piecewise smooth curve from $\exp_p(x)$ to $\exp_p(y)$ such that $L(c) < d(\exp_p(x), \exp_p(y)) + \varepsilon$. Since the Hausdorff dimensions of the sets of all conjugate cut points of p and all nonnormal cut points of p are not greater than $\dim M - 2$ (see [Itoh and Tanaka 1998, Lemmas 2 and 3; Federer

1969]), we may assume that c does not pass those points. Then we can apply Lemma 4.7 to obtain a pb-curve $\tilde{c}$ such that $\exp_p \circ \tilde{c} = c$. Note that if this pb-curve $\tilde{c}$ does not satisfy $\tilde{c}(a) = x$ and $\tilde{c}(b) = y$, then those end points are replaced by x and y, because $\lim_{t \to a} \exp_p(\tilde{c}(t)) = \exp_p(x)$ and $\lim_{t \to b} \exp_p(\tilde{c}(t)) = \exp_p(y)$. The resulting curve $\tilde{c}$ after this change is a pb-curve as well and connects from x to y such that $\exp_p(\tilde{c}(t)) = c(t)$ for all $t \in [a, b]$. Therefore, we have

$$d_0^*(x, y) \le L(\tilde{c}) = L(c) < d(\exp_p(x), \exp_p(y)) + \varepsilon,$$

and, hence, $d_0^*(x, y) \le d(\exp_p(x), \exp_p(y))$. □

We define an equivalence relation $\sim$ in $\widetilde{U}_p^c$ as follows: $x \sim y$ if and only if $d_0^*(x, y) = 0$. Let $[x]$ denote the equivalence class of this relation $\sim$ containing $x \in \widetilde{U}_p^c$ and $\widetilde{U}_p^c/\!\sim\, = \{[x] \mid x \in \widetilde{U}_p^c\}$. It follows from Lemma 5.1 that $[x] = [y]$ if and only if $\exp_p(x) = \exp_p(y)$ for any $x, y \in \widetilde{U}_p^c$. We define a metric $d_1^*([x], [y])$ on $\widetilde{U}_p^c/\!\sim$ by $d_1^*([x], [y]) = d_0^*(x, y)$ for any $x, y \in \widetilde{U}_p^c$.

Lemma 5.2. *Let d_1^* be the distance defined as above. Then (M, d) is isometric to $(\widetilde{U}_p^c/\!\sim, d_1^*)$ where d is the distance induced by F on M.*

The following theorem is a direct consequence of Lemma 5.2.

Theorem 5.3. *Let (M, F_M) and (N, F_N) be geodesically forward complete Finsler manifolds and $p_M \in M$, $p_N \in N$. Assume that there exists a linear isomorphism $I : T_{p_M}M \to T_{p_N}N$ such that $F_M(p_M, x) = F_N(p_N, I(x))$ for all $x \in T_{p_M}M$ and $\exp_{p_N}(I(x)) = \exp_{p_N}(I(y))$ for all $x, y \in \widetilde{C}(p_M)$ with $\exp_{p_M}(x) = \exp_{p_M}(y)$. If $(\widetilde{U}_{p_M}^c, d_M^*)$ and $(\widetilde{U}_{p_N}^c, d_N^*)$ are isometric under the map I, then (M, F_M) and (N, F_N) are isometric.*

Proof. From the assumption, we see that $\widetilde{C}(p_N) = I(\widetilde{C}(p_M))$ and $[I(x)] = [I(y)]$ for all $x, y \in \widetilde{C}(p_M)$ with $[x] = [y]$. Therefore, $(\widetilde{U}_{p_M}^c/\!\sim, {d_{p_M}}_1^*)$ is isometric to $(\widetilde{U}_{p_N}^c/\!\sim, {d_{p_N}}_1^*)$. This theorem follows from Lemma 5.2. □

Without assuming the invariance of the equivalence classes under the map I, we do not have an isometry from M to N. In fact, there exist surfaces (M, F_M) and (N, F_N) such that they are not isometric, although $(\widetilde{U}_{p_M}^c, d_M^*)$ and $(\widetilde{U}_{p_N}^c, d_N^*)$ are isometric.

Example 5.4. Let $a > b > 0$. Let T^2 be a torus defined by

$$T^2 := \big\{\big((a + b\cos u)\cos v, (a + b\cos u)\sin v, b\sin u\big) \bigm| 0 \le u \le 2\pi, 0 \le v \le 2\pi\big\}.$$

Let $M = T^2$. We make a surface N in the following way. Cut and open T^2 along the meridian circle $v = \pi$, and then glue the boundary as follows:

$$((a + b\cos u)\cos(\pi - 0), (a + b\cos u)\sin(\pi - 0), b\sin u)$$

and

$$((a + b\cos(-u))\cos(\pi + 0), (a + b\cos(-u))\sin(\pi + 0), b\sin(-u))$$

are identified for $-\pi \le u \le \pi$. The resulting surface N is a Klein bottle. Hence M is not isometric to N. However, if $p_M \in M$ and $p_N \in N$ are the points corresponding to $(a+b, 0, 0)$, then $(\widetilde{U}^c_{p_M}, d^*_M)$ is isometric to $(\widetilde{U}^c_{p_N}, d^*_N)$.

We next see the relation between $(\widetilde{U}^c_p, d^*)$ and (M, d). We say that a critical point $q \in C(p)$ is *jointing* if there exists a number $\delta > 0$ such that $B_b(q, \varepsilon) \cap B_f(p, d(p, q))$ is not connected for any $\varepsilon \in (0, \delta)$, where $B_b(x, \tau)$ and $B_f(x, \tau)$ are backward and forward open distance balls with center $x \in M$ and radius $\tau > 0$, respectively. It follows from Theorem 4.3 in [Innami et al. 2019] that if a nonconjugate cut point $q \in C(p)$ is a local minimum point of the distance function $d(p, \cdot)|_{C(p)}$, then q is jointing. Let $\mathrm{Join}(p)$ denote the set of all jointing cut points of p in M.

Lemma 5.5. *Let $q \in C(p)$ be jointing. Then $A_p(q)^s$ consists exactly two elements, say v and w, satisfying that $\omega_v + k\omega_w = 0$ for some number $k > 0$. In particular, q is a local minimum point of the function $d(p, \cdot)|_{C(p)}$.*

Proof. Since q is jointing, there exist at least two elements v and w in $A_p(q)^s$. Let $N(v) = \{z \in T_qM \mid \omega_v(z) = 0\}$. If $N(v) \ne N(w)$ for $v \ne w$, then there exists a curve $z(t)$, $t \in [0, 1]$, connecting $-v$ and $-w$ such that $\omega_v(z(t)) < 0$ and $\omega_w(z(t)) < 0$ for all $t \in [0, 1]$. This contradicts the fact that q is a jointing cut point. This implies that $\omega_v + k\omega_w = 0$ for some number $k > 0$ and there is no vector other than v and w in $A_p(q)^s$. □

Lemma 5.6. *Let $q \in C(p)$ be not jointing. Then, for any $\varepsilon > 0$, there exists a curve $c : [0, 1] \to M$ such that $c(0) = c(1) = q$, $d(p, c(t)) < d(p, q)$ for all $t \in (0, 1)$ and $L(c) < \varepsilon$.*

When $\dim M = 2$, we can see the detailed structure of the cut locus $C(p)$ in [Sabau and Tanaka 2016; Shiohama and Tanaka 1996]. Let $S_f(p, t)$ be the forward distance sphere with center p and radius t (the distance is measured from p) and $C_e(p)$ the set of all end cut points.

Lemma 5.7 [Shiohama and Tanaka 1996, Theorems A and B; Sabau and Tanaka 2016, Theorems B and C]. *Let* $\dim M = 2$. *Then the following are true.*

(1) *There exists a class $\mathscr{M} = \{m_1, \ldots\}$ of countably many rectifiable Jordan arcs $m_i : I_i \to C(p)$, $i = 1, \ldots,$ such that I_i is an open or closed interval and such that*

$$C(p) \setminus C_e(p) = \bigcup_{i=1}^{\infty} m_i(I_i), \quad \textit{disjoint union}$$

where each m_i has at most countably many branch points such that there are at most countably many members in $\mathscr{M}$ emanating from each of them.

(2) *There exists a set* $\mathscr{E}_M \subset (0, \infty)$ *of measure zero with the following properties. For every* $t \notin \mathscr{E}_M$ *with* $t > 0$, *there exist at most two minimal geodesics from* p *to every point* $x \in S_f(p, t) \cap C(p)$. *Furthermore, if* $x \in S_f(p, t) \cap C(p)$ *is joined from* p *by a unique minimal geodesic, then* x *is an end point of* $C(p)$. *There exists at most countably many points in* $S_f(p, t) \cap C(p)$ *which are joined from* p *by two distinct minimal geodesics.*

Some claims of this lemma can be translated into the tangent space T_pM under the exponential map $\exp_p$ as follows.

Lemma 5.8. *Let* $\dim M = 2$. *There exists a set* $\mathscr{E}_M \subset (0, \infty)$ *of measure zero with the following properties. For every* $t \notin \mathscr{E}_M$ *with* $t > 0$, *for every point* $x \in S_f(p, t) \cap C(p)$, *we have that* ${\exp_p}^{-1}(x)$ *consists of at most two points. Furthermore, if* ${\exp_p}^{-1}(x)$ *consists of a unique point, then* x *is an end point of* $C(p)$. *There exists at most countably many pairs of points* $\tilde{x}$ *and* $\tilde{y}$ *in* $\widetilde{S}(0, t) \cap \widetilde{C}(p)$ *such that* $\exp_p(\tilde{x}) = \exp_p(\tilde{y}) =: z$, *and, moreover, we may assume that* $\tilde{x}$ *and* $\tilde{y}$ *are interior points of* ${\exp_p}^{-1}(m_i)$ *for some members* m_i *in* $\mathscr{M}$. *In particular, there exist subarcs* $\ell_{\tilde{x}}$ *and* $\ell_{\tilde{y}}$ *of* ${\exp_p}^{-1}(m_i)$ *containing* $\tilde{x}$ *and* $\tilde{y}$, *respectively, with* $\exp_p(\ell_{\tilde{x}}) = \exp_p(\ell_{\ell_{\tilde{y}}}) \subset m_i$.

From the construction of m_i, we remark that, for every i, $d_i : I_i \to \mathbb{R}$ defined by $d_i(t) = d(p, m_i(t))$ for all $t \in I_i$ is strictly monotone.

Theorem 5.9. *Let* (M, F_M) *and* (N, F_N) *be geodesically forward complete Finsler orientable surfaces, that is,* $\dim M = 2$, *and* $p_M \in M$, $p_N \in N$. *Assume that* $d(p_M, \cdot)^{-1}(a) \cap \mathrm{Join}(p_M)$ *has at most one element for all* $a \in \mathbb{R}$. *If* $(\widetilde{U}^c_{p_M}, d^*_M)$ *and* $(\widetilde{U}^c_{p_N}, d^*_N)$ *are isometric under some linear map* I *from* $T_{p_M}M$ *to* $T_{p_N}N$, *then* M *and* N *are isometric.*

Proof. Let $\psi : M \setminus C(p_M) \to N \setminus C(p_N)$ be given by $\psi(x) = \exp_{p_N} \circ I \circ {\exp_{p_M}}^{-1}(x)$ for all $x \in M \setminus C(p_M)$. From Lemma 5.1, $B_f(p_M, r)$ is isometric to $B_f(p_N, r)$ for a sufficiently small $r > 0$. Let $R = \sup\{r \mid B_f(p_M, r) \text{ is isometric to } B_f(p_N, r)\}$. We prove $R = \infty$. Suppose for indirect proof that $R < \infty$.

We prove that the map ψ can be extended to $B_f(p_M, R) \cup S_f(p_M, R)$ isometrically. It follows from Lemma 5.6 that $\exp_{p_M}(\tilde{x}) = \exp_{p_M}(\tilde{y})$ if and only if $\exp_{p_N}(I(\tilde{x})) = \exp_{p_N}(I(\tilde{y}))$ for all $\tilde{x}, \tilde{y} \in S(0, R) \cap \widetilde{C}(p_M)$ which are not tangent jointing cut points. If $\exp_{p_M}(\tilde{x}) = \exp_{p_M}(\tilde{y}) =: x$ is jointing, then x is a local minimum point of the function $d(p_M, \cdot)|_{C(p_M)}$ because of Lemma 5.5. From the assumptions on $\mathrm{Join}(p_M)$, $(\widetilde{U}^c_{p_M}, d^*_M)$ and $(\widetilde{U}^c_{p_N}, d^*_N)$, we see that $\exp_{p_N}(I(\tilde{x}))$ is a local minimum point of $d(p_N, \cdot)|_{C(p_N)}$ and, moreover,

$$d(p_N, \cdot)^{-1}(d_M(p, x)) \cap \mathrm{Join}(p_N)$$

has exactly one point. Thus, from Lemma 5.5, $\exp_{p_N}(I(\tilde{x})) = \exp_{p_N}(I(\tilde{y})) =: x_N$, since $A_{p_N}(x_N)^s$ consists of exactly two elements.

Since we suppose $R < \infty$, there exists a sequence of numbers $t_j \notin \mathscr{E}_M \cap \mathscr{E}_N$, $t_j > R$, such that there exist points $\tilde{x}_j$ and $\tilde{y}_j$ with $\exp_{p_M}(\tilde{x}_j) = \exp_{p_M}(\tilde{y}_j) =: x_j \in S(p_M, t_j) \cap C(p_M)$ but $\exp_{p_N}(I(\tilde{x}_j)) \neq \exp_{p_N}(I(\tilde{y}_j))$ for every j and they converge to points $\tilde{x}$ and $\tilde{y}$ with $\exp_{p_M}(\tilde{x}) = \exp_{p_M}(\tilde{y}) =: x \in S(p_M, R) \cap C(p_M)$, respectively. However this cannot happen. In fact, we may assume that x_j are contained in one member m_x in $\mathscr{M}_M$ emanating from $x \in C(p_M)$. Let $\ell_{\tilde{x}}$ and $\ell_{\tilde{y}}$ be subarcs of $\widetilde{C}(p_M)$ containing all $\tilde{x}_j$ and $\tilde{y}_j$, respectively. Since $t_j \notin \mathscr{E}_M \cap \mathscr{E}_N$ and the functions $t \mapsto d(p, I(\ell_{\tilde{x}}(t)))$ and $t \mapsto d(p, I(\ell_{\tilde{y}}(t)))$ are strictly monotone, we see that $I(\ell_{\tilde{x}})$ and $I(\ell_{\tilde{y}})$ are identified and members in $\mathscr{M}_N$. This implies that $\exp_{p_N}(I(\tilde{x}_j)) = \exp_{p_N}(I(\tilde{y}_j))$, a contradiction. □

6. Differentiable points of the cut locus

Let (M, F) be a geodesically forward complete Finsler manifold without boundary. Let $P(v, w)$ denote the vector subspace of T_qM spanned by $v, w \in T_qM$. Note that if $\{v, w\}$ is linearly dependent and $w \neq 0$, then the dimension of $P(v, w)$ is one.

Lemma 6.1. *Let $q \in C(p)$ and $w \in T_qM \setminus T_qC(p)$. Then there exists a tangent vector $v \in T_qC(p)$ such that $P(v, w) \cap A_p(q) \setminus \{0\} \neq \emptyset$.*

Proof. Let $c(s)$, $s \in [0, \delta)$, be a smooth curve such that $c(0) = q$, $c'(0) = w$ and $c(s) \notin C(p)$ for all $s \in (0, \delta)$. Let $\gamma_s : [0, \ell_s] \to M$ be maximal minimal geodesics from p through $c(s)$. Then γ_s converges a minimal geodesic $\gamma : [0, d(p, q)] \to M$ from p to q because of Theorem 4.6 and $\gamma_s(\ell_s) \in C(p)$ converges to q as $s \to 0$. If v_s denotes the initial tangent vector of the unit speed minimal geodesics from q to $\gamma_s(\ell_s)$ at q, then there exists a subsequence v_{s_j} converging to a vector v and $\gamma'(d(p, q)) \in P(v, w) \cap A_p(q)$. □

Lemma 6.2. *Let X, Y and Z be vector spaces such that X is the direct sum of Y and Z. For $z_1 \in Z$ (resp. $z_2 \in Z$), let Z_1 (resp. Z_2) be the vector subspace spanned by Y and z_1 (resp. z_2) in X. Then either $Z_1 = Z_2$ or $Z_1 \cap Z_2 = Y$. Furthermore, if $\{z_1, z_2\}$ is linearly independent, we then have $Z_1 \cap Z_2 = Y$.*

Proof. We prove that $\{z_1, z_2\}$ is linearly dependent if $Z_1 \cap Z_2 \neq Y$. Suppose $Z_1 \cap Z_2 \neq Y$. Then there exists a vector $z \in Z_1 \cap Z_2 \setminus Y$. Hence $z = ay_1 + bz_1 = cy_2 + dz_2$ for some $y_1, y_2 \in Y$ and $a, b, c, d \in \mathbb{R}$ with $b \neq 0$, $d \neq 0$. Since $ay_1 - cy_2 = -bz_1 + dz_2$ and $Y \cap Z = \{0\}$, we have $z_2 = (b/d)z_1$. Therefore $Z_1 \cap Z_2 \neq Y$ implies that $Z_1 = Z_2$. □

Let $H(T_qC(p))$ and $L(T_qC(p))$ be the convex hull of $T_qC(p)$ in T_qM and the vector subspace generated by $H(T_qC(p))$ in T_qM, respectively. Obviously, $T_qC(p) \subset H(T_qC(p)) \subset L(T_qC(p))$. We say that $q \in C(p)$ is an *end cut point* of $C(p)$ if $\gamma'(d(p, q)) \in T_qC(p)$ for some minimal geodesic γ from p to q.

Lemma 6.3. *Let $q \in C(p)$ be not an end cut point and let a vector subspace V be a complement of $L(T_qC(p))$ in T_qM. Suppose* $\dim V \geq 1$. *Then $A_p(q)$ contains a cone which is a graph over V in $T_qM = V + L(T_qC(p))$. Namely, there exists a function $f : V \to T_qC(p)$ such that f is positively homogeneous and $A_p(q) = \{(w, f(w)) \mid w \in V\}$.*

Proof. Let $w \in V \setminus \{0\}$. As was seen in the proof of Lemma 6.1, we have a tangent vector $v \in T_qC(p)$ and a minimal geodesic segment γ from p to q such that $\gamma'(d(p,q)) \in P(v, w) \cap A_p(q)$ i.e., $\gamma'(d(p,q)) = aw + bv$ for certain numbers a, $b \in \mathbb{R}$. Since q is not an end cut point of $C(p)$, we have $a \neq 0$. Moreover, from Lemma 6.2, a map from $V \setminus \{0\}$ to T_qM, $aw \mapsto aw + bv$, is injective. In fact, if $aw_1 + bv_1 = cw_2 + dv_2$ for some $a, b, c, d \in \mathbb{R}$ with $a \neq 0$, $c \neq 0$ and $w_1, w_2 \in V$, $v_1, v_2 \in T_qC(p)$, then $\mathrm{Span}(w_1, L(T_qC(p))) = \mathrm{Span}(w_2, L(T_qC(p))) \neq L(T_qC(p))$. Therefore w_1 and w_2 are linearly dependent and, moreover, $aw_1 = cw_2$. Thus, we can define a map $f : V \to T_qC(p)$ by $f(w) = (b/a)v$ and $f(0) = 0$; then $A_p(q)$ is the graph of f over V in T_qM. From the construction of f, we see that f is positively homogeneous on V. □

Remark 6.4. Because there are various selection methods of the complement of $L(T_qC(p))$ in T_qM, this cone may not be determined uniquely (see Examples 7.2 and 7.4).

We say that $C(p)$ is *differentiable* at $q \in C(p)$ if $T_qC(p)$ is a vector subspace of T_qM, i.e., $L(T_qC(p)) = T_qC(p)$. From the definition, $C(p)$ is not differentiable at any end cut point of $C(p)$, because there exists a minimal geodesic γ from p to q such that $\gamma'(d(p,q)) \in T_qC(p)$ and $-\gamma'(d(p,q)) \notin T_qC(p)$, implying that $T_qC(p)$ is not a vector subspace.

The following theorem is a direct consequence of Lemma 6.3.

Theorem 6.5. *If $C(p)$ is differentiable at $q \in C(p)$, then $A_p(q)$ contains a cone which is a graph over the complement V of $T_qC(p)$ in T_qM. In particular, the union S of minimal geodesics $T(p,q)$ from p to q contains a sphere with dimension* $\dim V = n - \dim T_qC(p)$.

Corollary 6.6. *If $C(p)$ is differentiable at $q \in C(p)$, there is an $n - \dim T_qC(p) - 1$ sphere S in $\widetilde{C}(p)$ such that $\exp_p(T(0, \tilde{q}))$ is a minimal geodesic from p to q in M for the line segment $T(0, \tilde{q})$ from origin and ending any point $\tilde{q} \in S$.*

We define a distance function from p restricted to $C(p)$ by $d_{C(p)}(q) = d(p,q)$ for any $q \in C(p)$ and a distance function on M to $C(p)$ by ${d_{C(p)}}^b(x) = d(x, C(p))$ for any point $x \in M$. Let $q \in C(p)$. We say that a minimal geodesic $T(x,q)$ from x to q is a *perpendicular* to $C(p)$ with *foot* q if $d(y,q) = {d_{C(p)}}^b(y)$ for all $y \in T(x,q)$.

Corollary 6.7. *Let $q \in C(p)$ be a local minimum point of $d_{C(p)}$. Then there exists a number $\delta > 0$ such that $\gamma_v : [-\delta, 0] \to M$ is a perpendicular to $C(p)$ with foot q for any $v \in A_p(q)^s$. In particular, if $C(p)$ is differentiable at q, then the set of all perpendiculars to q is homeomorphic to a disk containing 0 in the complement V of $T_q C(p)$ for sufficiently small $\varepsilon > 0$.*

Proof. Let $\delta > 0$ be a number such that, for all $x \in B_b(q, 2\delta) \cap C(p)$, we have $d_{C(p)}(q) \le d_{C(p)}(x)$. For $v \in A_p(q)^s$, if $\gamma_v : [-d(p, q), 0] \to M$ is a geodesic with $\gamma_v{}'(0) = v$, then γ_v is a minimal geodesic from p to q. Let $p_1 = \gamma_v(-\delta)$. Then q is a foot of p_1 on $C(p)$. In fact, otherwise, there exists a point $q' \in B_b(q, 2\delta) \cap C(p)$ such that $d_{C(p)}(q') < d_{C(p)}(q)$, contradicting the choice of δ. This proves the corollary. □

7. Examples

It is well known that the cut loci of the compact rank one symmetric spaces (two point homogeneous spaces) are smooth. We construct other cut loci in product spaces such that they have differentiable points.

Example 7.1. Let $(M \times N, g = \alpha \times \beta)$ be the Riemannian product of two complete Riemannian manifolds (M, α) and (N, β). Let $p = (p_1, p_2) \in M \times N$. Then the cut locus $C(p)$ of p in $M \times N$ is given by $C(p) = C_M(p_1) \times N \cup M \times C_N(p_2)$ where $C_M(p_1)$ (resp. $C_N(p_2)$) is the cut locus of p_1 (resp. p_2) in M (resp. N). This follows from the fact: $\gamma(t) = (\mu(t), \nu(t))$ is a minimal geodesic from p to $q = (q_1, q_2)$ with unit speed for $t \in [0, d(p, q)]$ if and only if $\mu(t)$ (resp. $\nu(t)$) is the minimal geodesic from p_1 (resp. p_2) to q_1 (resp. q_2) with speed $d_M(p_1, q_1)/d(p, q)$ (resp. $d_N(p_2, q_2)/d(p, q)$). Therefore, we see, at $q = (q_1, q_2) \in C(p)$,

$$T_q C(p) = \begin{cases} T_{q_1} C_M(p_1) + T_{q_2} N =: S_1 & \text{if } q_1 \in C_M(p_1) \text{ and } q_2 \notin C_N(p_2), \\ T_{q_1} M + T_{q_2} C_N(p_2) =: S_2 & \text{if } q_1 \notin C_M(p_1) \text{ and } q_2 \in C_N(p_2), \\ S_1 \cup S_2 & \text{if } q_1 \in C_M(p_1) \text{ and } q_2 \in C_N(p_2). \end{cases}$$

Assume that $C_M(p_1)$ and $C_N(p_2)$ are differentiable. Then, $C(p)$ is differentiable at $q = (q_1, q_2) \in C(p)$ if $q_1 \notin C_M(p_1)$ or $q_2 \notin C_N(p_2)$, since $S_1 \cup S_2$ is a vector subspace only when one of S_1 and S_2 is the empty set, i.e., one of $C_M(p_1)$ and $C_N(p_2)$ is the empty set.

Let h be a smooth function on $M \times N$ such that $\|dh(v)\|_g < \|v\|_g$ for all $v \in T(M \times N)$. Then we have a Finsler metric F by $F(v) = \|v\|_g + dh(v)$ for all $v \in T(M \times N)$ which is a Randers change by an exact 1-form dh. The cut locus of p in $(M \times N, F)$ is the same as $(M \times N, g)$ (see [Innami et al. 2019]) and, hence, is smooth at the same points.

To be seen in the following example, the differentiability of a cut locus $C(p)$ in our sense is different from that of the distance function from p restricted to $C(p)$.

An example is like the cross shape of a thick long surface of revolution and a thin long surface of revolution.

Example 7.2. Let M_0 be a smooth surface of revolution whose Riemannian metric is given by

$$ds^2 = dt^2 + g(t)^2 d\theta^2, \quad 0 \le t \le \ell, \quad 0 \le \theta \le 2\pi,$$

where $g(t) > 0$ for all $t \in (0, \ell)$, $g(0) = g(\ell) = 0$, $g'(0) = -g'(\ell) = 1$. Let p and q be the south and north pole of M_0, i.e., the point with $t = 0$ and $t = \ell$, respectively. Then we have $C_{M_0}(p) = \{q\}$ and $A_p(q) = T_q M_0 \setminus \{0\}$. Assume that g is constant on some interval $[a, b] \subset (0, \ell)$. Let $B(x, r)$ and $B(y, r)$ be open distance balls around x and y, respectively, where $x = \bigl(\frac{1}{2}(a+b), 0\bigr)$ and $y = \bigl(\frac{1}{2}(a+b), \pi\bigr)$ are their coordinates.

Since $[a, b] \times [0, 2\pi]$ is a flat cylinder, for a sufficiently small $r > 0$, $B(x, r)$ and $B(y, r)$ are developed isometrically into the Euclid plane. The following surface S of revolution replaces $B(x, r)$ and $B(y, r)$ in M_0.

$$ds^2 = dt^2 + h(t)^2 d\varphi^2, \quad 0 \le t \le \ell_1, \quad 0 \le \varphi \le 2\pi,$$

where $h(t) > 0$ for all $t \in (0, \ell_1]$, $h(0) = 0$, $h'(0) = 1$ and $h(t) = r - (\ell_1 - t)$ for all $t \in (\ell_1 - \delta, \ell_1]$, for some $\delta > 0$. Let M be a surface such that two surfaces S_x and S_y isometric to S are glued to $M_0 \setminus B(x, r) \cup B(y, r)$ along each connected component of its boundary. Then $q \in C_M(p)$ and $C_M(p)$ is a smooth curve passing through q in a neighborhood of q. Furthermore, $C_M(p)$ near q is a geodesic through q because of symmetry of M with respect to meridians $\theta = \pi/2$ and $\theta = 3\pi/2$.

We see that $A_p(q)$ consists of two connected components with interior points. In fact, the set of all minimal geodesics from p to q consists of meridians defined by $\theta \in [-\theta_0, \theta_0] \cup [\pi - \theta_0, \pi + \theta_0]$ for some $\theta_0 > 0$. This implies that the differential of the distance function $d_p|_{C_M(p)}$ from p on $C_M(p)$ at q is not 0 and it is 0 for all initial tangent vectors of meridians with $\theta \in (-\theta_0, \theta_0) \cup (\pi - \theta_0, \pi + \theta_0)$. Therefore, $C_M(p)$ is differentiable in our sense but the distance function $d_p|_{C_M(p)}$ is not differentiable at the tangent vectors of the meridians with $\theta = -\theta_0, \theta_0, \pi - \theta_0, \pi + \theta_0$ (see Figure 1).

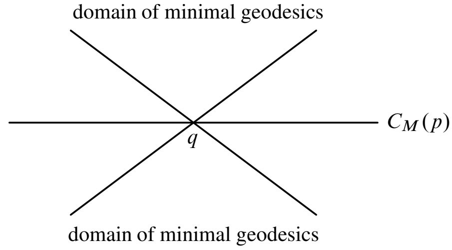


Figure 1. $C_M(p)$ around q.

We introduce the polar coordinate (r, ψ) in $T_q M$ such that $\psi = 0$ contains the tangent vector of $C_M(p)$ at q and $A_p(q)$ is the cone $\psi \in [\psi_0, \pi - \psi_0] \cup [\pi + \psi_0, 2\pi - \psi_0]$. If we choose the vector subspace $V = \{u(\cos\psi, \sin\psi) \mid u \in \mathbb{R}\}$ for any $\psi \in [\psi_0, \pi - \psi_0] \cup [\pi + \psi_0, 2\pi - \psi_0]$ as a complement to $T_q C_M(p)$, then V itself is a graph over V contained in $A_p(q)$ and obtained with the method in Lemma 6.1. While, if $V = \{u(\cos\psi, \sin\psi) \mid u \in \mathbb{R}\}$ for $\psi \in (0, \psi_0)$, then $\{u(\cos\psi_0, \sin\psi_0) \mid u \in \mathbb{R}\} \subset A_p(q)$ is the graph over V mentioned in Lemma 6.1.

Using the following Weinstein's result, we may make many examples of cut points as above.

Proposition 7.3 [Weinstein 1968, Proposition C]. *Let D be an n-disk embedded in a Riemannian manifold M^n. Then there is a new metric on M agreeing with the original metric on a neighborhood of M-(interior of D) such that, for some point p in D, $\exp_p$ is a diffeomorphism of the unit disk about the origin in $T_p M$ onto D.*

Example 7.4. Let S be a sphere having a flat domain Q. We first draw the smooth simple closed curve K in Q such that the cut locus $C(K)$ of K in the inside of K is like Figure 1. For example, let E be an ellipse in Q. We modify slightly arcs C_1 and C_2 near the end points of short axis of E in such a way that C_1 and C_2 are pieces of a circle with center at the center of the ellipse. We may assume that the resulting simple closed curve K is still symmetric with respect to the long axis and its cut locus $C(K)$ is like Figure 1. Then, we change the metric on the outside of K and find a point $p \in S$ as stated in Proposition 7.3. We have $C(p) = C(K)$ which is like Figure 1.

We check the exponential map and its differential map of the unit sphere in the Euclid space.

Example 7.5. Let (r, θ) be a geodesic polar coordinate about the north pole $p = (0, 0, 1)$ of the unit sphere

$$M = \{(x, y, z) \mid x^2 + y^2 + z^2 = 1\},$$

namely,

$$x = \sin r \cos\theta, \quad y = \sin r \sin\theta, \quad z = \cos r.$$

Let $D = \{(x, y) \mid x^2 + y^2 < 1\}$ be the unit disk. We use the orthogonal projection to xy-plane as a local coordinate system about the south half sphere of M. Then the exponential map $\exp_p$ at p is given by

$$\exp_p(r, \theta) = (\sin r \cos\theta, \sin r \sin\theta), \quad \frac{\pi}{2} < r < \frac{3\pi}{2}.$$

We see that $\exp_p(\pi, \theta) = q$ where $q = (0, 0, -1)$ is the south pole of M and q is the point conjugate to p along all minimal geodesics from p to q, since

$$d \exp_p |_{(r,\theta)} = \begin{pmatrix} \cos r \cos\theta & -\sin r \sin\theta \\ \cos r \sin\theta & \sin r \cos\theta \end{pmatrix}$$

and, hence,

$$d \exp_p |_{(\pi,\theta)} = \begin{pmatrix} -\cos\theta & 0 \\ -\sin\theta & 0 \end{pmatrix}.$$

Let $\tilde{c}(t) = (r(t), \theta(t))$ be a curve in T_pM. Then we have

$$c(t) = \exp_p(\tilde{c}(t)) = (\sin r(t) \cos\theta(t), \sin r(t) \sin\theta(t)).$$

If $c'(t) = (x'(t), y'(t))$, we then have the equation

$$\begin{cases} x'(t) = r'(t) \cos r(t) \cos\theta(t) - \theta'(t) \sin r(t) \sin\theta(t), \\ y'(t) = r'(t) \cos r(t) \sin\theta(t) + \theta'(t) \sin r(t) \cos\theta(t) \end{cases}$$

and, hence,

$$\begin{pmatrix} \cos r(t) \cos\theta(t) & -\sin r(t) \sin\theta(t) \\ \cos r(t) \sin\theta(t) & \sin r(t) \cos\theta(t) \end{pmatrix} \begin{pmatrix} r'(t) \\ \theta'(t) \end{pmatrix} = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix},$$

or

$$\begin{pmatrix} \cos\theta(t) & -\sin\theta(t) \\ \sin\theta(t) & \cos\theta(t) \end{pmatrix} \begin{pmatrix} \cos r(t) & 0 \\ 0 & \sin r(t) \end{pmatrix} \begin{pmatrix} r'(t) \\ \theta'(t) \end{pmatrix} = \begin{pmatrix} x'(t) \\ y'(t) \end{pmatrix}.$$

Hence, we have

$$r'(t) = \frac{1}{\cos r(t)} \big(x'(t) \cos\theta(t) + y'(t) \sin\theta(t) \big),$$
$$\theta'(t) = \frac{1}{\sin r(t)} \big(y'(t) \cos\theta(t) - x'(t) \sin\theta(t) \big).$$

Since $x^2 + y^2 = \sin^2 r$, $\cos\theta = x/\sqrt{x^2+y^2}$ and $\sin\theta = y/\sqrt{x^2+y^2}$, we have

$$r'(t) = \frac{x'(t)x(t) + y'(t)y(t)}{\sqrt{x(t)^2 + y(t)^2} \cos r(t)}, \qquad \theta'(t) = \frac{y'(t)x(t) - x'(t)y(t)}{x(t)^2 + y(t)^2}.$$

If $c(t) = (t, 0)$, then $\tilde{c}(t) = (\arcsin t, 0)$ for $t \in [0, 1]$. We consider a curve $c(t) = (t, bt^a)$, $t \in [0, 1]$, for $a > 1$ and $b \neq 0$. Then we have $c'(0) = (1, 0)$ and $\tilde{c}(0) = (\pi, 0)$.

Hence, $\cos r(t) < 0$ near $t = 0$. For $\tilde{c}'(0)$, since

$$r'(t) = -\frac{t + ab^2t^{2a-1}}{\sqrt{t^2 + b^2t^{2a}}\sqrt{1 - (t^2 + b^2t^{2a})}}, \qquad \theta'(t) = \frac{abt^a - bt^a}{t^2 + b^2t^{2a}},$$

we have $r'(t) \to -1$ and

$$\theta'(t) \to \begin{cases} \infty & \text{for } 1 < a < 2, \\ b & \text{for } a = 2, \\ 0 & \text{for } a > 2, \end{cases}$$

as $t \to 0+0$. Therefore, $\tilde{c}$ is differentiable at $t = 0$ if $a \geq 2$. Otherwise, $\tilde{c}$ cannot be extended differentiably to $t = 0$. This example shows us that we are not allowed to express $d\exp_p(\tilde{c}'(0)) = c'(0)$ in Lemma 4.4 for $1 < a < 2$ and $b \neq 0$.

References

[Berger 1960] M. Berger, "Sur quelques variétés riemanniennes suffisamment pincées", *Bull. Soc. Math. France* **88** (1960), 57–71. MR Zbl

[Berger 1961] M. Berger, "Sur les variétés à courbure positive de diamètre minimum", *Comment. Math. Helv.* **35** (1961), 28–34. MR Zbl

[Castelpietra and Rifford 2010] M. Castelpietra and L. Rifford, "Regularity properties of the distance functions to conjugate and cut loci for viscosity solutions of Hamilton–Jacobi equations and applications in Riemannian geometry", *ESAIM Control Optim. Calc. Var.* **16**:3 (2010), 695–718. MR Zbl

[Federer 1969] H. Federer, *Geometric measure theory*, Die Grundlehren der mathematischen Wissenschaften **153**, Springer, 1969. MR Zbl

[Figalli et al. 2011] A. Figalli, L. Rifford, and C. Villani, "Tangent cut loci on surfaces", *Differential Geom. Appl.* **29**:2 (2011), 154–159. MR Zbl

[Gibson 2001] C. G. Gibson, *Elementary geometry of differentiable curves*, Cambridge University Press, 2001. MR Zbl

[Gluck and Singer 1978] H. Gluck and D. Singer, "Scattering of geodesic fields, I", *Ann. of Math.* (2) **108**:2 (1978), 347–372. MR Zbl

[Hartman 1964] P. Hartman, "Geodesic parallel coordinates in the large", *Amer. J. Math.* **86** (1964), 705–727. MR Zbl

[Hebda 1983] J. J. Hebda, "The local homology of cut loci in Riemannian manifolds", *Tohoku Math. J.* (2) **35**:1 (1983), 45–52. MR Zbl

[Hebda 1994] J. J. Hebda, "Metric structure of cut loci in surfaces and Ambrose's problem", *J. Differential Geom.* **40**:3 (1994), 621–642. MR Zbl

[Innami et al. 2012] N. Innami, K. Shiohama, and T. Soga, "The cut loci, conjugate loci and poles in a complete Riemannian manifold", *Geom. Funct. Anal.* **22**:5 (2012), 1400–1406. MR Zbl

[Innami et al. 2019] N. Innami, Y. Itokawa, T. Nagano, and K. Shiohama, "Blaschke Finsler manifolds and actions of projective Randers changes on cut loci", *Trans. Amer. Math. Soc.* **371**:10 (2019), 7433–7450. MR Zbl

[Itoh 1996] J.-i. Itoh, "The length of a cut locus on a surface and Ambrose's problem", *J. Differential Geom.* **43**:3 (1996), 642–651. MR Zbl

[Itoh and Sabau 2016] J.-i. Itoh and S. V. Sabau, "Riemannian and Finslerian spheres with fractal cut loci", *Differential Geom. Appl.* **49** (2016), 43–64. MR Zbl

[Itoh and Sakai 2007] J.-i. Itoh and T. Sakai, "Cut loci and distance functions", *Math. J. Okayama Univ.* **49** (2007), 65–92. MR Zbl

[Itoh and Tanaka 1998] J.-i. Itoh and M. Tanaka, "The dimension of a cut locus on a smooth Riemannian manifold", *Tohoku Math. J.* (2) **50**:4 (1998), 571–575. MR Zbl

[Itoh and Tanaka 2001a] J.-i. Itoh and M. Tanaka, "The Lipschitz continuity of the distance function to the cut locus", *Trans. Amer. Math. Soc.* **353**:1 (2001), 21–40. MR Zbl

[Itoh and Tanaka 2001b] J.-i. Itoh and M. Tanaka, "A Sard theorem for the distance function", *Math. Ann.* **320**:1 (2001), 1–10. MR Zbl

[Klingenberg 1959] W. Klingenberg, "Contributions to Riemannian geometry in the large", *Ann. of Math.* (2) **69** (1959), 654–666. MR Zbl

[Myers 1935] S. B. Myers, "Connections between differential geometry and topology, I: Simply connected surfaces", *Duke Math. J.* **1**:3 (1935), 376–391. MR Zbl

[Nakagawa and Shiohama 1970a] H. Nakagawa and K. Shiohama, "On Riemannian manifolds with certain cut loci", *Tohoku Math. J.* (2) **22** (1970), 14–23. MR

[Nakagawa and Shiohama 1970b] H. Nakagawa and K. Shiohama, "On Riemannian manifolds with certain cut loci, II", *Tohoku Math. J.* (2) **22** (1970), 357–361. MR Zbl

[Omori 1968] H. Omori, "A class of Riemannian metrics on a manifold", *J. Differential Geometry* **2** (1968), 233–252. MR Zbl

[Ozols 1974] V. Ozols, "Cut loci in Riemannian manifolds", *Tohoku Math. J.* (2) **26** (1974), 219–227. MR Zbl

[Ozols 1976] V. Ozols, "Largest normal neighborhoods", *Proc. Amer. Math. Soc.* **61**:1 (1976), 99–101. MR Zbl

[Rifford 2004] L. Rifford, "A Morse–Sard theorem for the distance function on Riemannian manifolds", *Manuscripta Math.* **113**:2 (2004), 251–265. MR Zbl

[Sabau and Tanaka 2016] S. V. Sabau and M. Tanaka, "The cut locus and distance function from a closed subset of a Finsler manifold", *Houston J. Math.* **42**:4 (2016), 1157–1197. MR Zbl

[Shen 2001] Z. Shen, *Lectures on Finsler geometry*, World Scientific Publishing Co., Singapore, 2001. MR Zbl

[Shiohama and Tanaka 1996] K. Shiohama and M. Tanaka, "Cut loci and distance spheres on Alexandrov surfaces", pp. 531–559 in *Actes de la Table Ronde de Géométrie Différentielle* (Luminy, 1992), edited by A. L. Besse, Sémin. Congr. **1**, Soc. Math. France, 1996. MR Zbl

[Shiohama et al. 2003] K. Shiohama, T. Shioya, and M. Tanaka, *The geometry of total curvature on complete open surfaces*, Cambridge Tracts in Mathematics **159**, Cambridge University Press, 2003. MR Zbl

[Tanaka 2003] M. Tanaka, "Characterization of a differentiable point of the distance function to the cut locus", *J. Math. Soc. Japan* **55**:1 (2003), 231–241. MR Zbl

[Warner 1965] F. W. Warner, "The conjugate locus of a Riemannian manifold", *Amer. J. Math.* **87** (1965), 575–604. MR Zbl

[Warner 1967] F. W. Warner, "Conjugate loci of constant order", *Ann. of Math.* (2) **86** (1967), 192–212. MR Zbl

[Warner 1971] F. W. Warner, *Foundations of differentiable manifolds and Lie groups*, Scott, Foresman and Co., Glenview, Il, 1971. MR Zbl

[Weinstein 1968] A. D. Weinstein, "The cut locus and conjugate locus of a riemannian manifold", *Ann. of Math.* (2) **87** (1968), 29–41. MR Zbl

Received November 4, 2019.

NOBUHIRO INNAMI
DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE
NIIGATA UNIVERSITY
NIIGATA
JAPAN
innami@math.sc.niigata-u.ac.jp

YOE ITOKAWA
DEPARTMENT OF INFORMATION AND COMMUNICATION ENGINEERING
FUKUOKA INSTITUTE OF TECHNOLOGY
WAJIRO-HIGASHI
FUKUOKA
JAPAN
itokawa@fit.ac.jp

TOSHIKI KONDO
GRADUATE SCHOOL OF SCIENCE AND TECHNOLOGY
NIIGATA UNIVERSITY
NIIGATA
JAPAN
f17j003a@mail.cc.niigata-u.ac.jp

TETSUYA NAGANO
DEPARTMENT OF INFORMATION SECURITY
UNIVERSITY OF NAGASAKI
NAGASAKI
JAPAN
hnagano@sun.ac.jp

KATSUHIRO SHIOHAMA
FUKUOKA INSTITUTE OF TECHNOLOGY
WAJIRO-HIGASHI
FUKUOKA
JAPAN
k-siohama@fit.ac.jp

PACIFIC JOURNAL OF MATHEMATICS
Vol. 308, No. 1, 2020
https://doi.org/10.2140/pjm.2020.308.103

SHIFT OPERATORS, RESIDUE FAMILIES AND DEGENERATE LAPLACIANS

ANDREAS JUHL AND BENT ØRSTED

In this paper, we introduce new aspects in conformal geometry of some very natural second-order differential operators. These operators are termed shift operators. In the flat space, they are intertwining operators which are closely related to symmetry breaking differential operators. In the curved case, they are closely connected with ideas of holography and the works of Fefferman–Graham, Gover–Waldron and one of the authors. In particular, we obtain an alternative description of the so-called residue families in conformal geometry in terms of compositions of shift operators. This relation allows easy new proofs of some of their basic properties. In addition, we derive new holographic formulas for Q-curvatures in even dimension. Since these turn out to be equivalent to earlier holographic formulas, the novelty here is their conceptually very natural proof. The overall discussion leads to a unification of constructions in representation theory and conformal geometry.

1. Introduction and formulation of the main results

Conformal differential geometry has seen spectacular developments in recent years, both from a perspective of pure mathematics, and from a mathematical physics point of view. The construction of ambient metrics and Poincaré metrics by Fefferman and Graham [2012] gave rise to many important applications and fundamental insights. This is closely connected to the idea of holography.

MSC2010: primary 35J30, 53A30, 53B20; secondary 35Q76, 53C25, 58J50.
Keywords: Poincaré metrics, ambient metrics, conformal geometry, symmetry breaking operators, residue families, shift operators, GJMS operators, Q-curvature.

Attempts to extend these ideas to a conformal submanifold theory were one source for the notion of symmetry breaking operators. This recent notion in representation theory is central in studies of, for instance, the interplay between the representation theory of the conformal group (the Möbius group) of Euclidean space and the corresponding group for a hyperplane. In particular, it plays a basic role in the study of branching laws of representations. The works [Kobayashi et al. 2015; 2016; Kobayashi and Speh 2015; 2018; Kobayashi and Pevzner 2016; Fischmann et al. 2016; Frahm and Ørsted 2019] reflect recent progress in this area. Curved analogs of symmetry breaking operators in conformal geometry deal with conformally covariant differential operators $C^\infty(X) \to C^\infty(M)$, where M is a hypersurface of a Riemannian manifold X (see [Juhl 2009]). Residue families (introduced in [Juhl 2009]) are curved versions of symmetry breaking operators which are defined in a setting where X is a tubular neighborhood of M and where the metric on X is determined by the metric on M. For recent substantial progress in the general case we refer to [Gover and Peterson 2018].

In this paper, we shall develop a theory of some second-order differential operators, originally found via representation theory as *shift operators* between symmetry breaking operators. These operators turn out to be very natural and admit generalizations within a framework defined by Riemannian metrics. Remarkably, these generalizations did appear in earlier work from a number of different perspectives, in particular from the point of view of tractor calculus, a powerful tool in conformal geometry.

Shift operators recently appeared in the theory of symmetry breaking operators. The latter operators are generalizations of Knapp–Stein intertwining operators which intertwine principal series representations of semisimple Lie group. Symmetry breaking operators map between functional spaces on a given flag variety to functional spaces on a subvariety and are equivariant only with respect to the symmetry group of the subvariety. This loss of symmetry is the origin of the notion. A typical situation is that of the round sphere S^{n+1} with an equatorially embedded subsphere $S^n \hookrightarrow S^{n+1}$. The noncompact model of that situation is a standard embedding $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+1}$. In these cases, the relevant groups are the conformal groups of the respective submanifolds. Conformal symmetry breaking *differential* operators acting on functions in that setting are of particular importance for conformal differential geometry [Juhl 2009].

Shift operators shift the spectral parameter in the distributional Schwartz kernels of conformal symmetry breaking operators. Such results in the setting $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+1}$ first appeared in [Fischmann et al. 2019] and will be recalled in Section 2E. The basic shift operator in that theory is given by the 1-parameter family [Fischmann et al. 2019, (3.5)]

$$P(\lambda) = r\Delta - (2\lambda - n - 3)\partial_r : C^\infty(\mathbb{R}^{n+1}) \to C^\infty(\mathbb{R}^{n+1}), \quad \lambda \in \mathbb{C} \tag{1-1}$$

of second-order differential operators on $\mathbb{R}^{n+1}$. Here Δ denotes the nonpositive Laplacian of the flat metric on the space $\mathbb{R}^{n+1}$ with coordinates (r, x). We regard r as a defining function of the subspace $\mathbb{R}^n$. For any $\lambda \in \mathbb{C}$, the operator $P(\lambda)$ is equivariant with respect to principal series representations restricted to the conformal group $\mathrm{Conf}(\mathbb{R}^n)$ of the subspace $\mathbb{R}^n$ with the flat metric regarded as the subgroup of the conformal group $\mathrm{Conf}(\mathbb{R}^{n+1})$ of $\mathbb{R}^{n+1}$ leaving $\mathbb{R}^n$ invariant. More precisely, $P(\lambda)$ satisfies the intertwining relation

$$\left(\frac{\gamma_*(r)}{r}\right)^{n-\lambda+2} \circ \gamma_* \circ P(\lambda) = P(\lambda) \circ \left(\frac{\gamma_*(r)}{r}\right)^{n-\lambda+1} \circ \gamma_* \tag{1-2}$$

for all $\gamma \in \mathrm{Conf}(\mathbb{R}^n) \subset \mathrm{Conf}(\mathbb{R}^{n+1})$. Here $\gamma_* = (\gamma^{-1})^*$ denotes the push-forward operator on functions induced by γ.

In the present paper, we generalize these results to a framework defined by Riemannian metrics and discuss some applications. One of the main applications concerns the residue families of [Juhl 2009]. As noted above, they can be regarded as curved analogs of symmetry breaking differential operators. More precisely, residue families are 1-parameter families of conformally covariant differential operators

$$D_N^{\mathrm{res}}(h; \lambda) : C^\infty(M_+) \to C^\infty(M), \quad \lambda \in \mathbb{C} \tag{1-3}$$

of order $N \in \mathbb{N}$. These are defined in the following setting. We consider a general Riemannian manifold (M, h). Let M_+ be an open neighborhood of $\{0\} \times M$ in $[0, \infty) \times M$. On the open interior $M_+^\circ = (0, \varepsilon) \times M$ of M_+, let $g_+ = r^{-2}(dr^2 + h_r)$ be an even Poincaré metric in normal form relative to h. Here h_r is a 1-parameter family of metrics on M with $h_0 = h$. The metric $\bar{g} = dr^2 + h_r$ is a conformal compactification of g_+. The relevant concepts were developed in [Fefferman and Graham 2012] and will be recalled in Section 2. Although the Poincaré metric g_+ is not completely determined by the metric h, residue families only depend on the Taylor coefficients of the family h_r at $r = 0$ which are uniquely determined by h. More precisely, the even-order family $D_{2N}^{\mathrm{res}}(h; \lambda)$ involves $2N$ derivatives by the variable r and depends on the Taylor coefficients of h_r of order $\leq 2N$. The conformal covariance of residue families describes their behavior under conformal changes $h \to e^{2\varphi} h$ of the metric on the submanifold $M \hookrightarrow M_+ = [0, \varepsilon) \times M$.

Our generalizations of the shift operator $P(\lambda)$ are differential operators which act on smooth functions on M_+ and are defined in terms of an even Poincaré metric g_+ on M_+°. In fact, we define a curved version of the shift operator $P(\lambda)$ by the formula[1]

$$S(g_+; \lambda) = r\Delta_{\bar{g}} - (2\lambda - n + 1)\partial_r - \tfrac{1}{2}(\lambda - n + 1)\,\mathrm{tr}(h_r^{-1}\dot{h}_r). \tag{1-4}$$

[1] The shift by 2 in the parameter λ is a matter of conventions.

Here the dot denotes derivatives with respect to r and $\bar{g} = r^2 g_+$. This definition can also be written in the form

$$S(g_+; \lambda) = r\Delta_{\bar{g}} - (2\lambda - n+1)\partial_r - (\lambda - n+1)\dot{v}(r)/v(r),$$

where the function $v(r, \cdot) \in C^\infty(M)$ is defined by the relation

$$d\mathrm{vol}(h_r) = v(r)\, d\mathrm{vol}(h)$$

of volume forms. We note that, in contrast to residue families, the shift operators are not defined only by the Taylor coefficients of h_r at $r = 0$.[2]

The operator $S(g_+; \lambda)$ is a second-order differential operator which degenerates for $r = 0$, i.e., on the submanifold M. Theorem 3.7 establishes the shift property of $S(g_+; \lambda)$. This property describes its action on functions of the form $r^\lambda u \in C^\infty(M_+^\circ)$, where u is an eigenfunction of the Laplacian Δ_{g_+} of the Poincaré metric g_+ on M_+°.

The following result describes the behavior of $S(g_+; \lambda)$ under conformal changes of the boundary metric h (Proposition 3.9).

Theorem 1. *Assume that (M^n, h) is a manifold of dimension n. Let $\hat{h} = e^{2\varphi} h$ be a metric in the conformal class of h. Let g_+ be an even Poincaré metric in normal form relative to h on M_+°. Let κ be a diffeomorphism of M_+ which restricts to the identity on M and for which the Poincaré metric $\hat{g}_+ = \kappa^*(g_+)$ is in normal form relative to $\hat{h}$. In these terms, we have*

$$S(\hat{g}_+; \lambda) = \kappa^* \circ \left(\frac{\kappa_*(r)}{r}\right)^{\lambda - n} \circ S(g_+; \lambda) \circ \left(\frac{\kappa_*(r)}{r}\right)^{n-\lambda-1} \circ \kappa_*. \tag{1-5}$$

This result may be regarded as a version of conformal covariance. Although the transformation law (1-5) formally resembles the equivariance property (1-2), the former law is *not* a generalization of the latter one. In fact, the diffeomorphisms κ should not be confused with the conformal maps γ: κ leaves the submanifold M pointwise fixed. However, the formal similarity between the intertwining property (1-2) and the conformal transformation law (1-5) can be explained by recognizing $P(\lambda)$ and $S(g_+; \lambda)$ (for Einstein g_+) both as special cases of the degenerate Laplacian $I{\cdot}D$ introduced in [Gover and Waldron 2014] (this concept will be recalled in Section 2G). Indeed, we note that

$$P(\lambda) = S(g_{\mathrm{hyp}}; \lambda - 2)$$

and

$$S(g_{\mathrm{hyp}}; \lambda) = (I{\cdot}D)[r^2 g_{\mathrm{hyp}}; r, \lambda - n+1] \quad \text{and} \quad S(g_+; \lambda) = -(I{\cdot}D)[r^2 g_+; r, \lambda - n+1],$$

[2]However, in the analytic category, the family h_r is completely determined by h.

where g_{hyp} denotes the hyperbolic metric in the upper half-space and g_+ is Einstein (see (2-38) and (3-4)). Since γ preserves the hyperbolic metric g_{hyp}, we have

$$\gamma^*(r^2 g_{\mathrm{hyp}}) = \left(\frac{\gamma^*(r)}{r}\right)^2 (r^2 g_{\mathrm{hyp}})$$

and the conformal transformation law for $I{\cdot}D$ (Proposition 2.2) implies

$$\begin{aligned}
&\left(\frac{r}{\gamma^*(r)}\right)^{\lambda-n-2} \circ \gamma^* \circ P(\lambda) \circ \gamma_* \circ \left(\frac{r}{\gamma^*(r)}\right)^{n-\lambda+1} \\
&= \left(\frac{r}{\gamma^*(r)}\right)^{\lambda-n-2} \circ \gamma^* \circ (I{\cdot}D)[r^2 g_{\mathrm{hyp}}; r; \lambda-n-1] \circ \gamma^* \circ \left(\frac{r}{\gamma^*(r)}\right)^{n-\lambda+1} \\
&= \left(\frac{r}{\gamma^*(r)}\right)^{\lambda-n-2} \circ \gamma^* \circ (I{\cdot}D)[\gamma^*(r^2 g_{\mathrm{hyp}}); \gamma^*(r); \lambda-n-1] \circ \gamma^* \circ \left(\frac{r}{\gamma^*(r)}\right)^{n-\lambda+1} \\
&= (I{\cdot}D)[r^2 g_{\mathrm{hyp}}; r, \lambda-n-1] = P(\lambda).
\end{aligned}$$

This proves (1-2). A similar calculation gives (1-5) (Remark 3.10).

For $N \in \mathbb{N}$, we define the compositions

$$S_N(g_+;\lambda) \stackrel{\text{def}}{=} \underbrace{S(g_+;\lambda) \circ \cdots \circ S(g_+;\lambda+N-1)}_{N \text{ factors}}. \tag{1-6}$$

We shall refer to these operators as *iterated shift operators* or simply also as shift operators. Theorem 1 implies that all iterated shift operators $S_N(g_+;\lambda)$ are conformally covariant (in the sense as in (1-5)). The following result states that residue families (1-3) can be written in terms of iterated shift operators (Corollary 4.2). Its proof rests on the shift property of the shift operators. Let the embedding $\iota : M \hookrightarrow M_+$ be defined by $m \mapsto (0, m)$.

Theorem 2. *Assume that (M^n, h) is a Riemannian manifold of dimension n. Let $N \in \mathbb{N}$ so that $2N \leq n$ if n is even. Then the residue family $D_{2N}^{\mathrm{res}}(h;\lambda)$ of order $2N$ is proportional to the composition of the family $\lambda \mapsto S_{2N}(g_+;\lambda+n-2N)$ with the restriction ι^* to M. More precisely, we have*

$$(-2N)_N \left(\lambda + \frac{n+1}{2} - 2N\right)_N D_{2N}^{\mathrm{res}}(h;\lambda) = \iota^* S_{2N}(g_+;\lambda+n-2N). \tag{1-7}$$

A similar formula holds true for odd-order residue families.

Some comments on this result are in order.

By construction, the family $S_{2N}(g_+;\lambda)$ involves $4N$ derivatives in the variable r and depends on all Taylor coefficients of h_r. The identity (1-7) shows that its composition with the restriction operator ι^* actually involves only $2N$ derivatives in r and depends only on the Taylor coefficients of h_r up to order $2N$. In particular, the compositions $\iota^* S_{2N}(g_+;\lambda)$ are completely determined by h. In the following, we shall denote these compositions by $\Sigma_{2N}(h;\lambda)$. For $N \in \mathbb{N}$ not satisfying the assumptions in Theorem 2, the compositions $\iota^* S_{2N}(g_+;\lambda)$ in general are not determined only by h.

The product formula (1-7) yields a new expression for the residue families. It extends a result of [Clerc 2017; Fischmann et al. 2019] in the flat case. The description of residue families in terms of restrictions to M of "powers" of a universal shift operator living in a neighborhood of M resembles the construction of the conformally covariant powers of the Laplacian (GJMS operators) of h by powers of the Laplacian of an ambient metric associated to h [Graham et al. 1992].

Theorem 2 can be used to deduce properties of residue families from properties of shift operators and vice versa. In particular, the conformal covariance of $S_N(g_+;\lambda)$ implies a conformal covariance law for residue families. This reproves [Juhl 2009, Theorem 6.6.3].

In addition, Theorem 2 enables us to give easy proofs of the systems of factorization identities of residue families which play an important role in [Juhl 2009; Juhl 2013] in connection with the description of recursive structures for GJMS operators and Q-curvatures. Our new proofs of these factorization identities rest on two basic facts. The first one (Theorem 5.1) is also of independent interest.

Theorem 3. *Assume that $N \in \mathbb{N}$ with $2N \leq n$ if n is even. Then*

$$S_N\Big(g_+;\frac{n-1}{2}\Big) = r^N P_{2N}(\bar{g}), \tag{1-8}$$

up to an error term in $O(r^\infty)$ for n odd and $o(r^{n-N})$ for n even. Moreover, the equality holds true without an error term if g_+ is Einstein.

Here $P_{2N}(\bar{g})$ is a GJMS operator of the conformal compactification $\bar{g}$ of the Poincaré metric g_+ in normal form relative to h.[3] The second basic fact is the identity

$$\Sigma_{2N}\Big(h;\frac{n}{2}-N\Big) = ((2N-1)!!)^2 P_{2N}(h)\iota^* \tag{1-9}$$

(Theorem 5.6). This formula reproves a special case of a result of [Gover and Waldron 2014]. Together with

$$\Sigma_{2N}\Big(h;\frac{n-1}{2}-N\Big) = (2N)!\iota^* P_{2N}(\bar{g}) \tag{1-10}$$

it shows that the operators $\Sigma_{2N}(h;\lambda)$ interpolate between GJMS operators for the metrics h and $\bar{g}$.

The coefficients of the families $S_N(g_+;\lambda)$ depend on the parameters r and λ. A closer study of both dependencies seems to be of interest. Theorem 3 may be

[3]For odd n, the operators $P_{2N}(\bar{g})$ are well-defined for all $N \in \mathbb{N}$. See the comments at the beginning of Section 5A.

regarded as a result in that direction. More results in this direction are discussed in Section 6F.

Finally, through the relation between residue families and iterated shift operators, we derive a new formula for the critical Q-curvature $Q_n(h)$ of a manifold (M^n, h) of even dimension n (Theorem 5.12).

Theorem 4. *Let n be even. Then*

$$Q_n(h) = c_n \Sigma_{n-1}(h; 0)\partial_r(\log v), \tag{1-11}$$

where $c_n = (-1)^{n/2} 2^{n-2} \big(\Gamma\big(\tfrac{n}{2}\big)/\,\Gamma(n)\big)^2$.

There is an interesting formal resemblance of the latter formula for the critical Q-curvature with a formula of Fefferman and Hirachi [2003].

Theorem 4 extends to all subcritical Q-curvatures $Q_{2N}(h)$ for $2N < n$ in the form

$$Q_{2N}(h) = c_{2N} \Sigma_{2N-1}\Big(h; \frac{n}{2} - N\Big)\partial_r(\log v), \tag{1-12}$$

where $c_{2N} = (-1)^N 2^{2N-2} (\Gamma(N)/\,\Gamma(2N))^2$ (Theorem 5.14).

Combining (1-11) and (1-12) with Theorem 2, yields formulas for Q-curvatures in terms of residue families. These turn out to be equivalent to the holographic formulas proved in [Graham and Juhl 2007; Juhl 2011]. In other words, these holographic formulas for Q-curvatures can be viewed as natural consequences of Theorem 2.

The operator $S(g_+; \lambda)$ appeared in the literature in different contexts. The construction of asymptotic expansions of eigenfunctions for the Laplacian Δ_{g_+} of a Poincaré metric in [Graham and Zworski 2003] involved a second-order operator $\mathcal{D}_s$. In [Gover and Waldron 2014], the operator $\mathcal{D}_s$ was interpreted and generalized within tractor calculus. This led to the definition of the so-called degenerate Laplacian $I{\cdot}D$ which played a role in the discussion after Theorem 1. Compositions as in (1-6) of these operators were used in [Gover and Waldron 2014] in connection with the construction of asymptotic expansions in a more general eigenfunction problem. The relations among these construction will be described in Section 2. In [Clerc 2017], Clerc gave a representation theoretical alternative construction of a family of symmetry breaking differential operators introduced in [Juhl 2009] in terms of compositions of shifted operators $P(\lambda)$. Theorem 2 is a generalization of his result to the curved setting.

The paper is organized as follows. After a collection of background material, we use Section 3 to introduce the shift operator $S(g_+; \lambda)$ and prove basic properties. In Section 4, we establish the connection between residue families and iterated shift operators. Section 5 is devoted to various applications. Here we provide easy new proofs of the factorization identities of residue families and discuss holographic formulas for Q-curvatures. In Section 6, we illustrate the main results in low-order cases. In the final section, we speculate on the role of iterated shift operators in the theory of the building block operators $\mathcal{M}_{2N}$ [Juhl 2013; 2016] of GJMS operators. In this connection, we derive a new formula for the so-called holographic Laplacian [Juhl 2016] for the metric $\bar{g}$.

2. Preliminaries

In the present section, we fix notation and describe the general setting. We recall basic facts on GJMS operators, Q-curvatures, residue families, shift operators and the degenerate Laplacian.

2A. ***General notation.*** $\mathbb{N}$ is the set of natural numbers and $\mathbb{N}_0$ the set of nonnegative integers. For a complex number $a \in \mathbb{C}$ and an integer $N \in \mathbb{N}$, the Pochhammer symbol $(a)_N$ is defined by $(a)_N \stackrel{\text{def}}{=} a(a+1)\cdots(a+N-1)$. We also set $(a)_0 \stackrel{\text{def}}{=} 1$. $C^\infty(M)$ is the space of smooth functions on the manifold M and $C_c^\infty(M)$ denotes the subspace of functions with compact support. The Laplacian of a Riemannian metric g on a manifold M is denoted Δ_g; it acts on $C^\infty(M)$. Here we use the convention that $-\Delta_g$ is nonnegative, i.e., $-\Delta_g = \delta_g d$, where δ_g is the formal adjoint of the differential d. $\mathrm{Ric}(g)$ and $\tau(g)$ denote the Ricci tensor and the scalar curvature of g. On a manifold (M^n, g) of dimension n, we set $\mathsf{J}(g) = \frac{1}{2(n-1)}\tau(g)$ and define the Schouten tensor of g by $\mathsf{P}(g) = \frac{1}{n-2}(\mathrm{Ric}(g) - \mathsf{J}(g)g)$. We shall also write simply P and J if the metric is clear by context. The symbol $\circ$ denotes compositions of operators.

2B. ***GJMS operators and Q-curvatures.*** Let (M^n, h) be a Riemannian manifold of dimension $n \geq 3$. For $N \in \mathbb{N}$ if n is odd and $\mathbb{N} \ni N \leq \frac{n}{2}$ if n is even, the GJMS operators are conformally covariant differential operators

$$P_{2N}(h) : C^\infty(M) \to C^\infty(M)$$

of order $2N$ which are of the form $P_{2N}(h) = \Delta_h^N + \mathrm{LOT}$, where LOT denotes lower-order terms. These lower-order terms only depend on covariant derivatives of the curvature of h. Under conformal changes $\hat{h} = e^{2\varphi}h$ with $\varphi \in C^\infty(M)$ of the metric, the GJMS operators satisfy

$$P_{2N}(\hat{h}) = e^{-(n/2+N)\varphi} \circ P_{2N}(h) \circ e^{(n/2-N)\varphi}. \tag{2-1}$$

In [Graham et al. 1992], these operators were constructed in terms of powers of the Laplacian of an ambient metric associated to h.

The GJMS operators generalize the well-known Yamabe operator

$$P_2 = \Delta - \left(\frac{n}{2} - 1\right)\mathsf{J} \tag{2-2}$$

and the Paneitz operator

$$P_4 = \Delta^2 + \delta((n-2)\mathsf{J}h - 4\mathsf{P})\#d + \left(\frac{n}{2}-2\right)\left(\frac{n}{2}\mathsf{J}^2 - 2|\mathsf{P}|^2 - \Delta\mathsf{J}\right), \tag{2-3}$$

where $|\mathsf{P}|^2 = \mathsf{P}_{ij}\mathsf{P}^{ij}$ and # indicates the natural action of symmetric 2-tensors on $\Omega^1(M)$.

In odd dimension n, we have GJMS operators P_{2N} of any order $2N$, $N \in \mathbb{N}$. But, for general metrics h in even dimension n, the restriction $2N \le n$ is necessary both for the definition of $P_{2N}(h)$ and for the existence of conformally covariant differential operators with leading term Δ_h^N [Graham 1992; Gover and Hirachi 2004].

Explicit formulas for GJMS operators for general metrics are very complicated [Juhl 2013]. But for some special metrics, they may be given by closed formulas. In particular, for Einstein manifolds (M^n, h) they are given by the formula

$$P_{2N}(h) = \prod_{l=1}^{N}\left(\Delta_h - 2\mu\left(\frac{n}{2}+l-1\right)\left(\frac{n}{2}-l\right)\right) \tag{2-4}$$

for all $N \in \mathbb{N}$, where the constant $\mu \in \mathbb{R}$ is defined by $\mathrm{Ric}(h) = 2\mu(n-1)h$ [Fefferman and Graham 2012]. Here the above restriction on their order is irrelevant.

It is a basic observation [Branson 1995] that

$$P_{2N}(h)(1) = (-1)^N\left(\frac{n}{2}-N\right)Q_{2N}(h) \tag{2-5}$$

for a scalar curvature invariant $Q_{2N}(h) \in C^\infty(M^n)$ of order $2N$. The quantities $Q_{2N}(h)$ are well-defined by (2-5) as long as $2N < n$. These curvature quantities are called the subcritical Q-curvatures. Their analogs of even order n can be defined by analytic continuation in dimension n through the subcritical Q-curvatures. The quantity $Q_n(h)$ is called the critical Q-curvature of (M^n, h). Under the respective conditions $n > 2$ and $n > 4$, (2-2) and (2-3) yield the subcritical Q-curvatures

$$Q_2 = \mathsf{J} \quad \text{and} \quad Q_4 = \frac{n}{2}\mathsf{J}^2 - 2|\mathsf{P}|^2 - \Delta\mathsf{J} \tag{2-6}$$

of order 2 and 4. Here we suppress the obvious dependence of constructions on the metric h. By continuation in dimension n, we define the respective critical Q-curvatures

$$Q_2 = \mathsf{J} \quad \text{and} \quad Q_4 = 2\mathsf{J}^2 - 2|\mathsf{P}|^2 - \Delta\mathsf{J} \tag{2-7}$$

in dimension $n = 2$ and $n = 4$.

In the following, we shall often simplify notation by omitting the composition sign $\circ$ in compositions with multiplication operators. It also will often lead to simplifications to suppress the obvious dependence of constructions on h.

2C. ***Poincaré metrics, eigenfunction expansions and GJMS operators.*** In the present section, we briefly recall basic definitions concerning Poincaré metrics in the sense of Fefferman and Graham [2012] and recall a description of GJMS operators of (M, h) in terms of eigenfunctions of the Laplacian of an associated Poincaré metric on M_+° [Graham and Zworski 2003]. This description will be of central importance for all later constructions.

Let M be a manifold of dimension $n \geq 3$. Let M_+ be an open neighborhood of $\{0\} \times M$ in $[0, \infty) \times M$, i.e., $M_+ = [0, \varepsilon) \times M$ for some $\varepsilon > 0$. We use the coordinate r on the first factor. We define the embedding $\iota : M \to M_+$ by $\iota(m) = (0, m)$. Let $M_+^\circ = (0, \varepsilon) \times M$. A smooth metric

$$g_+ = r^{-2}(dr^2 + h_r) \tag{2-8}$$

on M_+° is called a Poincaré metric in normal form relative to a metric h on M if $\bar{g} = r^2 g_+$ extends to M_+, $\bar{g}$ restricts to h, i.e., $\iota^*(\bar{g}) = h$, and the Ricci tensor of g_+ satisfies the Einstein condition

$$\mathrm{Ric}(g_+) + n g_+ = O(r^\infty) \tag{2-9}$$

for odd $n \geq 3$ and the Einstein condition

$$\mathrm{Ric}(g_+) + n g_+ = O(r^{n-2}) \tag{2-10}$$

together with the vanishing trace condition

$$\mathrm{tr}_h\big(\iota^*(r^{-n+2}(\mathrm{Ric}(g_+) + n g_+))\big) = 0 \tag{2-11}$$

for even $n \geq 4$. The metric $\bar{g} = r^2 g_+$ on M_+ is called a conformal compactification of g_+. The family h_r in (2-8) is a smooth 1-parameter family of metrics on M.

If, for odd n, we also assume that h_r has an *even* expansion

$$h_r = h_0 + r^2 h_2 + r^4 h_4 + \cdots, \tag{2-12}$$

then the condition (2-9) implies that the coefficients $h_2, h_4, \ldots$ are uniquely determined by $h_0 = h$. These metrics are conveniently referred to as *even* Poincaré metrics. One may consider the conformal compactification $\bar{g}$ of an even Poincaré metric as a smooth metric on the larger space $(-\varepsilon, \varepsilon) \times M$. In the real analytic category, h_r converges and $\mathrm{Ric}(g_+) + n g_+ = 0$ in a neighborhood of $\{0\} \times M$.

For even n, the situation is more complicated. In that case, the family h_r has an expansion of the form

$$h_r = \underbrace{h_0 + r^2 h_2 + \cdots + r^{n-2} h_{n-2}}_{\text{even powers}} + r^n (h_n + \log r\, h_0^{(1)}) + \cdots. \tag{2-13}$$

The condition (2-10) uniquely determines the coefficients $h_2, \dots, h_{n-2}$ by $h_0 = h$. Moreover, the vanishing trace condition (2-11) can be satisfied and determines the h-trace of h_n. However, the trace-free part of h_n is *not* determined by h.

In general, the higher-order solutions of the Einstein condition contain $\log r$ terms. The first $\log r$ coefficient $h_0^{(1)}$ (Fefferman–Graham obstruction tensor) is uniquely determined by h and trace-free. For specific choices of the trace-free part of h_n, the condition

$$\mathrm{Ric}(g_+) + n g_+ = O(r^\infty)$$

may be satisfied by solutions with expansions of the form

$$h_r = \underbrace{h_0 + r^2 h_2 + \cdots + r^n h_n + \cdots}_{\text{even powers}} + \sum_{j=1}^{\infty} (r^n \log r)^j h_r^{(j)} \tag{2-14}$$

with even families $h_r^{(j)}$. The $\log r$ terms in these expansion vanish if and only if the obstruction tensor vanishes.

Now assume that $g_+ = r^{-2}(dr^2 + h_r)$ is a Poincaré metric in normal form relative to h. Hence $h_0 = h$. Let $\hat{h} = e^{2\varphi} h$ be a metric in the same conformal class as h. Then a suitable change of coordinates brings g_+ into normal form relative to $\hat{h}$. Following [Graham and Lee 1991, Section 5; Fefferman and Graham 2012, Proposition 4.3], we briefly recall the arguments proving this basic observation. The metric g_+ is asymptotically hyperbolic since $|dr/r|_{g_+} = 1$ on $r = 0$. The latter property suffices to prove the existence of $u \in C^\infty(M_+)$ so that for $\rho = re^u \in C^\infty(M_+)$ we have

$$|d\rho|^2_{\rho^2 g_+} = 1 \quad \text{near } M.$$

Here the restriction of u to $r = 0$ can be arbitrarily chosen. Now let $\mathfrak{X} = \mathrm{grad}_{\rho^2 g_+}(\rho)$ be the gradient field of ρ with respect to the conformal compactification $\rho^2 g_+$ of g_+. Let $\Phi^t_{\mathfrak{X}}$ be the flow of $\mathfrak{X}$. In these terms, we define the map

$$\kappa : [0, \varepsilon) \times M \ni (\lambda, x) \mapsto (\Phi^\lambda_{\mathfrak{X}})(x) \in [0, \varepsilon) \times M$$

for sufficiently small ε. Then $\kappa(0, x) = x$ and $\kappa^*(\rho)(\lambda, x) = \lambda$. The gradient field $\mathfrak{X}$ is orthogonal to the slices $\rho^{-1}(\lambda)$. It follows that

$$\kappa^*(\rho^2 g_+) = d\lambda^2 + k_\lambda$$

for some 1-parameter family k_λ. Hence

$$\kappa^*(g_+) = \frac{1}{\kappa^*(\rho)} \kappa^*(\rho^2 g_+) = \lambda^{-2}(d\lambda^2 + k_\lambda).$$

Finally, we note that

$$k_0 = \iota^*(d\lambda^2 + k_\lambda) = \iota^* \kappa^*(\rho^2 g_+) = (\kappa\iota)^*(\rho^2 g_+) = \iota^*(\rho^2 g_+) = \iota^* \left(\frac{\rho}{r}\right)^2 h_0 = e^{2\iota^*(u)} h_0$$

(with obvious embeddings ι). In other words, for the choice $\iota^*(u) = \varphi$, $\kappa^*(g_+)$ is a Poincaré metric in normal form relative to $\hat{h}$.

The volume function $v(r) \in C^\infty(M^n)$ is defined by the relation

$$d\mathrm{vol}(h_r) = v(r)\, d\mathrm{vol}(h) \tag{2-15}$$

of volume forms. For odd n and even Poincaré metrics, the function $v(r)$ has an even Taylor series $v(r) = 1 + r^2 v_2 + r^4 v_4 + \cdots$. Similarly, for even n, we have

$$v(r) = 1 + r^2 v_2 + \cdots + r^n v_n + \cdots .$$

The indicated coefficients in these expansions are locally determined by h. The coefficients v_{2j} are called the renormalized volume coefficients of (M^n, h) [Graham 2000]. Note that

$$\dot{v}(r)/v(r) = \tfrac{1}{2}\,\mathrm{tr}(h_r^{-1}\dot{h}_r). \tag{2-16}$$

In particular, for even n, the coefficient v_n only depends on the trace of h_n. We also set

$$w(r) \overset{\text{def}}{=} \sqrt{v(r)}. \tag{2-17}$$

Then $w(r) = \sum_{j \geq 0} r^{2j} w_{2j}$. The coefficients w_{2N} are polynomials in v_{2k} for $k \leq N$.

A routine calculation shows that the Laplacian of g_+ takes the form

$$\Delta_{g_+} = r^2 \Delta_{h_r} + r^2 \partial_r^2 - (n-1) r \partial_r + \tfrac{1}{2}\,\mathrm{tr}(h_r^{-1}\dot{h}_r) r^2 \partial_r, \tag{2-18}$$

where Δ_{h_r} is the Laplacian of h_r and $\dot{h}_r$ denotes the derivative of h_r with respect to r. Since $\bar{g}$ and g_+ are conformally equivalent, another calculation shows that

$$\Delta_{\bar{g}} = r^{-2}(\Delta_{g_+} + (n-1) r \partial_r) = \Delta_{h_r} + \partial_r^2 + \tfrac{1}{2}\,\mathrm{tr}(h_r^{-1}\dot{h}_r)\partial_r. \tag{2-19}$$

For even h_r, this formula is well-defined on $(-\varepsilon, \varepsilon) \times M$.

We continue with the discussion of GJMS operators of (M^n, h). These operators can be described in terms of asymptotic expansions of the solutions of the equation

$$\Delta_{g_+} u + \lambda(n - \lambda) u = 0, \ \lambda \in \mathbb{C}.$$

In the case of the hyperbolic ball such solutions can be represented as Helgason–Poisson transforms of distributions (or even hyperfunctions) on the boundary S^n [Helgason 1984]. Here we consider solutions with *smooth* boundary values in the curved setting. A corresponding Poisson transform was constructed in [Graham and Zworski 2003, Proposition 3.5]. Its definition rests on a local (near the boundary) asymptotic analysis of the expansions of eigenfunctions and global mapping properties of the resolvent of Δ_{g_+} acting on appropriate functional spaces.

More precisely, we consider eigenfunctions u with asymptotic expansions of the form

$$u(r,x) \sim \sum_{j\geq 0} r^{\lambda+2j} a_{2j}(\lambda)(x) + \sum_{j\geq 0} r^{n-\lambda+2j} b_{2j}(\lambda)(x) \tag{2-20}$$

with coefficients $a_{2j}(\lambda), b_{2j}(\lambda) \in C^\infty(M)$. The coefficients in both sums in (2-20) are determined by the respective leading coefficients $a_0(\lambda)$ and $b_0(\lambda)$ through a recursive algorithm. Moreover, for a global eigenfunction u, both leading coefficients are related by a scattering operator $\mathcal{S}(\lambda)$.

We recall these constructions in some more detail. First, a local asymptotic analysis yields a map

$$\Phi(\lambda) : C^\infty(M) \to r^{n-\lambda} C^\infty(M_+), \quad \Re(\lambda) > n/2$$

so that

$$(\Delta_{g_+} + \lambda(n-\lambda))\Phi(\lambda) f = O(r^\infty).$$

For even Poincaré metrics, it has the form

$$\Phi(\lambda) f = r^{n-\lambda} f + \sum_{j\geq 1} r^{n-\lambda+2j} \mathcal{T}_{2j}(n-\lambda) f,$$

where $\mathcal{T}_{2j}(\lambda)$ are meromorphic families of differential operators on M of order $2j$. Next, the resolvent $R(\lambda) = (\Delta_{g_+} + \lambda(n-\lambda))^{-1} : L^2(M_+) \to L^2(M_+)$ is holomorphic for $\Re(\lambda) > n$. Moreover, its restriction to the space of smooth functions which vanish of infinite order on the boundary, admits a meromorphic continuation to $\mathbb{C}$. The range of that restriction of $R(\lambda)$ is contained in $r^\lambda C^\infty(M_+)$. Then the family

$$\mathcal{P}(\lambda) \stackrel{\text{def}}{=} \Phi(\lambda) - R(\lambda)(\Delta_{g_+} + \lambda(n-\lambda))\Phi(\lambda) \tag{2-21}$$

of *Poisson transforms* is meromorphic on $\Re(\lambda) > n/2$ with poles only for real λ with $\lambda(n-\lambda) \in \sigma_d(-\Delta_{g_+}) \subset (0, (n/2)^2)$.[4] $\mathcal{P}(\lambda)$ is continuous up to $\Re(\lambda) = n/2$, $\lambda \neq n/2$. It satisfies

$$(\Delta_{g_+} + \lambda(n-\lambda))\mathcal{P}(\lambda) = 0.$$

Away from the real poles in $\Re(\lambda) > n/2$ and $\lambda \notin n/2 + \mathbb{N}_0$, we have

$$\mathcal{P}(\lambda) f = r^\lambda G + r^{n-\lambda} F \tag{2-22}$$

for $F, G \in C^\infty(M_+)$ with $\iota^*(F) = f$.[5] f is viewed as the *boundary value* of the eigenfunction $u = \mathcal{P}(\lambda) f$. For the details see [Graham and Zworski 2003, Proposition 3.5]. Later we shall use (2-22) for $\Re(\lambda) = n/2$, $\lambda \neq n/2$.

[4]The poles of $\Phi(\lambda)$ in $n/2 + \mathbb{N}$ cancel in the sum (2-21).

[5]For $\lambda \in n/2 + \mathbb{N}$ the corresponding expansion contains a $\log r$-term.

The scattering operator $\mathcal{S}(\lambda) : C^\infty(M) \to C^\infty(M)$ is defined by

$$\mathcal{S}(\lambda) : f \mapsto \iota^*(G)$$

for G as in (2-22). It follows that in the asymptotic expansion (2-20) of $u = \mathcal{P}(\lambda) f$ the coefficients are given by

$$b_{2j}(\lambda) = \mathcal{T}_{2j}(n-\lambda) f \quad \text{and} \quad a_{2j}(\lambda) = \mathcal{T}_{2j}(\lambda)\mathcal{S}(\lambda) f.$$

Note that $\mathcal{T}_0(\lambda) = \mathrm{Id}$.

The families $\mathcal{T}_{2j}(\lambda)$ only depend on the Taylor series of h_r. For odd n and even Poincaré metrics, these are determined by h. Therefore, we write $\mathcal{T}_{2j}(\lambda) = \mathcal{T}_{2j}(h; \lambda)$. For even n, only the families $\mathcal{T}_{2j}(\lambda)$ with $2j \le n$ are determined by h and we indicate that dependence accordingly. However, the scattering operator $\mathcal{S}(\lambda)$ is a global object which depends on the chosen metric on M_+°.

The families $\mathcal{T}_{2j}(h; \lambda)$ are meromorphic in λ, with simple poles at $\lambda = \frac{n}{2} - k$ for $k = 1, \dots, j$. Under the restriction $2j \le n$ for even n, the residue of $\mathcal{T}_{2j}(h; \lambda)$ at $\lambda = \frac{n}{2} - j$ is proportional to the GJMS operator $P_{2j}(h)$ on (M, h). More precisely, we have the basic residue formula

$$\mathrm{Res}_{\lambda=n/2-j}(\mathcal{T}_{2j}(h; \lambda)) = \frac{1}{2^{2j} j!(j-1)!} P_{2j}(h) \tag{2-23}$$

describing GJMS operators in terms of asymptotic expansions of eigenfunctions of Δ_{g_+}. In [Graham and Zworski 2003, Section 4] this formula is derived from the original ambient metric definition of the GJMS operators.

The residue at $\lambda = \frac{n}{2} + N$ of the right-hand side of the expansion (2-20) yields the contribution

$$r^{n/2+N}\big(\mathrm{Res}_{n/2+N}(\mathcal{S}(\lambda)) + \mathrm{Res}_{n/2-N}(\mathcal{T}_{2N}(\lambda))\big).$$

Under mild assumptions, this residue vanishes. Hence (2-23) implies the residue formula

$$\mathrm{Res}_{\lambda=n/2+j}(\mathcal{S}(\lambda)) = -\frac{1}{2^{2j} j!(j-1)!} P_{2j}(h) \tag{2-24}$$

for the poles of the scattering operator [Graham and Zworski 2003, Theorem 1].

We emphasize that the formula (2-23) only rests on the local analysis of eigenfunctions near the boundary $r = 0$. However, the definition of the scattering operator and the construction of *exact* eigenfunctions involves the *global* resolvent $R(\lambda)$ on an asymptotically hyperbolic manifold M_+°. For a given closed M, a simple choice for M_+° is $M_+^\circ = (0, 1) \times M$. In that case, the metric on M_+° is a Poincaré metric near *both* boundary components at $r = 0$ and $r = 1$. The resulting scattering operator then acts on smooth functions on the disjoint union of both copies of M.

A simple special case of the latter situation is the scattering operator of the hyperbolic cylinder $M_+^\circ = \Gamma\backslash\mathbb{H}^{n+1}$ by a cocompact discrete subgroup of $\mathrm{SO}(1,n)^\circ$ regarded as a subgroup of $\mathrm{SO}(1,n+1)^\circ$ (using a trivial embedding). Then M_+° can be identified with a cylinder $(-\infty,\infty)\times M$ with compact cross-section $M = \Gamma\backslash\mathbb{H}^n$. The boundary consists of two copies of M. The scattering operator of the cylinder acts on $C^\infty(M)\oplus C^\infty(M)$. It decomposes into the direct sum of endomorphisms on the spaces $E(\mu)\oplus E(\mu)$ generated by the eigenspaces

$$E(\mu) = \bigl\{u \in C^\infty(M) \mid -\Delta u = \mu(n-1-\mu)u\bigr\},$$

where Δ is the Laplacian of the hyperbolic metric on M. The restriction $\mathcal{S}(\lambda;\mu)$ of $\mathcal{S}(\lambda)$ to this space is given by [Patterson and Perry 2001, Appendix B]

$$\mathcal{S}(\lambda;\mu) = 2^{n-2\lambda}\frac{1}{\pi}\frac{\Gamma\left(\frac{n}{2}-\lambda\right)}{\Gamma\left(\lambda-\frac{n}{2}\right)}\Gamma(\lambda-\mu)\Gamma(\lambda-(n-1-\mu))\begin{pmatrix}\sin\pi\left(\frac{n}{2}-\mu\right) & \sin\pi\left(\frac{n}{2}-\lambda\right)\\ \sin\pi\left(\frac{n}{2}-\lambda\right) & \sin\pi\left(\frac{n}{2}-\mu\right)\end{pmatrix}.$$

Although $\mathcal{S}(\lambda;\mu)$ contains off-diagonal terms, its residues at $\frac{n}{2}+N$ are diagonal. More precisely, we find

$$\mathrm{Res}_{\lambda=n/2+N}(\mathcal{S}(\lambda;\mu)) = -\frac{1}{2^{2N}N!(N-1)!}\prod_{j=n/2}^{n/2+N-1}\bigl(-\mu(n-1-\mu)+j(n-1-j)\bigr)\,\mathrm{Id}\,.$$

This result implies the residue formula

$$\begin{aligned}\mathrm{Res}_{\lambda=n/2+N}(\mathcal{S}(\lambda)) &= -\frac{1}{2^{2N}N!(N-1)!}\prod_{j=n/2}^{n/2+N-1}(\Delta_M+j(n-1-j))\\ &= -\frac{1}{2^{2N}N!(N-1)!}P_{2N}(M,g_{\mathrm{hyp}}),\end{aligned}$$

which confirms the residue formula (2-24) of [Graham and Zworski 2003].

2D. *Residue families.* We recall the concept of residue families introduced in [Juhl 2009]. We assume that g_+ is an even Poincaré metric relative to h. For $N\in\mathbb{N}_0$ with $2N\le n$ for even n, we define a polynomial 1-parameter family of differential operators $C^\infty(M_+)\to C^\infty(M)$ through

$$D_{2N}^{\mathrm{res}}(h;\nu) \stackrel{\mathrm{def}}{=} 2^{2N}N!\Bigl(-\frac{n}{2}-\nu+N\Bigr)_N\delta_{2N}(h;\nu+n-2N), \tag{2-25}$$

$$D_{2N+1}^{\mathrm{res}}(h;\nu) \stackrel{\mathrm{def}}{=} 2^{2N}N!\Bigl(-\frac{n}{2}-\nu+N+1\Bigr)_N\delta_{2N+1}(h;\nu+n-2N-1), \tag{2-26}$$

where the family $\delta_N(h,\nu): C^\infty(M_+)\to C^\infty(M)$ is defined by the residue formula

$$\mathrm{Res}_{\lambda=-\nu-1-N}\biggl(\int_{M_+} r^\lambda u\varphi\, d\mathrm{vol}(\bar g)\biggr) = \int_M f\delta_N(h;\nu)\varphi\, d\mathrm{vol}(h). \tag{2-27}$$

Here u is an eigenfunction of $-\Delta_{g_+}$ on M_+° with eigenvalue $\nu(n-\nu)$ and boundary value f (Section 2C), and we use test functions $\varphi \in C_c^\infty(M_+)$ ([Juhl 2009, (6.6.11)]). Note that $D_0^{\mathrm{res}}(\lambda) = \iota^*$. The above definitions yield the formula

$$\delta_N(h;\nu) = \sum_{j=0}^{N} \frac{1}{(N-j)!}\bigl[\mathcal{T}_j^*(h;\nu)v_0 + \cdots + \mathcal{T}_0^*(h;\nu)v_j\bigr]\iota^*\partial_r^{N-j} \tag{2-28}$$

([Juhl 2009, Definition 6.6.2]) in terms of *solution operators* $\mathcal{T}_{2j}(h;\nu)$ and renormalized volume coefficients v_{2j}. Here the operator $\mathcal{T}_j^*(h;\nu)$ denotes the formal adjoint of $\mathcal{T}_j(h;\nu)$ with respect to the scalar product on $C^\infty(M)$ defined by h. Since g_+ is assumed to be even, solution operators and renormalized volume coefficients with odd indices vanish. Although the residue families $D_N^{\mathrm{res}}(h;\nu)$ are defined in terms of asymptotic expansions of eigenfunctions of Δ_{g_+}, the assumptions on N guarantee that they only depend on the Taylor coefficients of h_r which are determined by h. This justifies the notation. For even n, the family $D_n^{\mathrm{res}}(h;\nu)$ sometimes is called the critical residue family. The formula (2-28) shows that, for even n, also the odd-order residue family $D_{n+1}^{\mathrm{res}}(h;\nu)$ is determined by h.

In later sections, we shall prefer to use the notation $D_N^{\mathrm{res}}(h;\lambda)$ instead of $D_N^{\mathrm{res}}(h;\nu)$.

2E. ***Shift operators and distributional kernels.*** We recall some results of [Fischmann et al. 2019]. We regard $\mathbb{R}^n$ as a subspace of $\mathbb{R}^{n+1}$ using the embedding $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+1}$ defined by $x \mapsto (0,x)$. Elements of $\mathbb{R}^{n+1}$ are written in the form (r,x). We regard $\mathbb{R}^{n+1}$ as a Riemannian manifold with the flat Euclidean metric g_0. The metric g_0 is the conformal compactification of the upper half-space $(\mathbb{R}_+^{n+1}, g_{\mathrm{hyp}})$ with the hyperbolic metric $g_{\mathrm{hyp}} \stackrel{\mathrm{def}}{=} r^{-2}g_0$. The hyperbolic metric g_{hyp} is a Poincaré metric g_+ in normal form relative to the flat metric h_0 on the boundary $\mathbb{R}^n$ of $\mathbb{R}_+^{n+1}$.

Let

$$K_{\lambda,\nu}^+(r,x) \stackrel{\mathrm{def}}{=} |r|^{\lambda+\nu-n-1}(|x|^2+r^2)^{-\nu} \quad \text{and} \quad K_{\lambda,\nu}^-(r,x) \stackrel{\mathrm{def}}{=} rK_{\lambda-1,\nu}^+(r,x) \tag{2-29}$$

be the distributional Schwartz kernels studied in [Kobayashi and Speh 2015; Frahm and Ørsted 2019]. The maps

$$\varphi \mapsto \int_{\mathbb{R}^{n+1}} K_{\lambda,\nu}^+(r,x-y)\varphi(r,x)\,dr\,dx$$

define operators $C_c^\infty(\mathbb{R}^{n+1}) \to C^\infty(\mathbb{R}^n)$ which are equivariant with respect to principal series representations of the conformal group of the subspace $\mathbb{R}^n \hookrightarrow \mathbb{R}^{n+1}$. Sometimes they are referred to as *symmetry breaking operators* [Kobayashi and Speh 2015]. Moreover, we set

$$P(\lambda) \stackrel{\mathrm{def}}{=} r\Delta - (2\lambda-n-3)\partial_r : C^\infty(\mathbb{R}^{n+1}) \to C^\infty(\mathbb{R}^{n+1}) \tag{2-30}$$

(see [Fischmann et al. 2019, (3.5)]), where Δ is the Laplacian of the flat metric g_0 on $\mathbb{R}^{n+1}$. The operator $P(\lambda)$ is equivariant with respect to principal series representations for the conformal group of $\mathbb{R}^n$ (see (1-2)). The following result motivates us to refer to this operator as a *shift operator* for the kernel $K^{\pm}_{\lambda,\nu}(r,x)$ [Fischmann et al. 2019, Theorem 3.5].

Proposition 2.1. *The operator $P(\lambda)$ shifts the λ-parameter of the kernels $K^{\pm}_{\lambda,\nu}(r,x)$, i.e.,*

$$P(\lambda)K^{\pm}_{\lambda,\nu}(r,x) = (\lambda+\nu-n-1)(\nu-\lambda+1)K^{\mp}_{\lambda-1,\nu}(r,x).$$

It was proven in [Fischmann et al. 2019, Theorem 4.2] that composition of shifted versions of $P(\lambda)$ recover the conformal symmetry breaking differential operators $D_N(\lambda)$ studied in [Juhl 2009; Kobayashi et al. 2015; Kobayashi and Speh 2015]. For a related representation theoretical proof of this fact we refer to [Clerc 2017].[6]

2F. *The operator $\mathcal{D}_\lambda(g_+)$.* Assume we are in the setting of Section 2B. In particular, (M^n, h) is a Riemannian manifold and g_+ is a Poincaré metric on M_+° in normal form relative to h. Following Graham and Zworski [2003, (4.4)], we introduce a differential operator $\mathcal{D}_\lambda(g_+)$ on $C^\infty(M_+^\circ)$ which plays a basic role in the construction of the solution operators $\mathcal{T}_{2j}(h;\lambda)$ (see Section 2C). This operator is defined by the equation

$$r^{-(n-\lambda+1)}[\Delta_{g_+} + \lambda(n-\lambda)]r^{n-\lambda} = -\mathcal{D}_\lambda(g_+), \quad \lambda \in \mathbb{C} \tag{2-31}$$

and has the explicit form

$$\mathcal{D}_\lambda(g_+) = -r\partial_r^2 + \Big(2\lambda-n-1-\frac{r}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r)\Big)\partial_r - \frac{n-\lambda}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r) - r\Delta_{h_r}. \tag{2-32}$$

In the case of the hyperbolic upper-half space $(\mathbb{R}^{n+1}_+, g_{\mathrm{hyp}})$, it is given by

$$\mathcal{D}_\lambda(g_{\mathrm{hyp}}) = -r\partial_r^2 + (2\lambda-n-1)\partial_r - r\Delta_{h_0} = -r\Delta_{g_0} + (2\lambda-n-1)\partial_r.$$

Comparing this formula with (2-30), yields the relation

$$P(\lambda) = -\mathcal{D}_{\lambda-1}(g_{\mathrm{hyp}}). \tag{2-33}$$

2G. *The degenerate Laplacian.* We recall the definition of the degenerate Laplacian introduced by Gover and Waldron [2014]. Let (X, c) be an $n+1$-dimensional conformal Riemannian manifold equipped with a scale $\sigma \in C^\infty(X)$. Associated to these data, we define the operator

$$u \mapsto -\sigma\Delta_g u + (n+2\omega-1)\Big[g(d\sigma, du) - \frac{\omega}{n+1}\Delta_g(\sigma)u\Big] - \frac{2\omega}{n+1}(n+\omega)\sigma\,\mathsf{J}(g)u \tag{2-34}$$

[6] The operator $P(\lambda)$ appears as (4.6) in [Clerc 2017].

on $C^\infty(X)$ (see [Gover and Waldron 2014, (2.9)]). Here g is a metric in the conformal class c and $\omega \in \mathbb{C}$. The above operator will be denoted by $(I{\cdot}D)[g,\sigma;\omega]$. The notation $I{\cdot}D$ for the degenerate Laplacian reflects its definition as a scalar product (in a tractor bundle) of a scale tractor I and the tractor operator D mapping functions to tractors. We shall not go here into the definitions of the relevant concepts of tractor calculus. Let $M = \sigma^{-1}(0)$ be the zero-locus of σ. We assume that σ is a defining function for the hypersurface M. Let $\iota : M \hookrightarrow X$ denote the embedding of M.

It follows from the definition that the operator $\iota^*(I{\cdot}D)[g,\sigma;\omega]$ degenerates to the first-order operator

$$u \mapsto (n+2\omega-1)\iota^*\Big[g(d\sigma, du) - \frac{\omega}{n+1}\Delta_g(\sigma)u\Big].$$

If $\sigma^{-2}g$ has constant scalar curvature $-n(n-1)$, it follows that $|\mathrm{grad}_g(\sigma)|^2 = 1$ on M and this operator reduces to the conformally covariant Robin type boundary operator

$$u \mapsto (n+2\omega-1)\iota^*(\nabla_{\mathrm{grad}_g(\sigma)} - \omega H_g)u,$$

where H_g is the mean curvature of M [Gover 2010, Section 3.1; Juhl 2009, Section 6.2].

By its very definition in terms of tractor calculus, the operator $I{\cdot}D$ satisfies a conformal covariance property. For the convenience of the reader, we provide an independent proof of that basic property.

Proposition 2.2. *The degenerate Laplacian satisfies*

$$(I{\cdot}D)[e^{2\varphi}g, e^{\varphi}\sigma;\omega] \circ e^{\omega\varphi} = e^{(\omega-1)\varphi} \circ (I{\cdot}D)[g,\sigma;\omega], \quad \omega \in \mathbb{C} \tag{2-35}$$

for all metrics g, scales $\sigma \in C^\infty(X)$ and $\varphi \in C^\infty(X)$. Moreover,

$$(I{\cdot}D)[\kappa^*(g), \kappa^*(\sigma);\omega] = \kappa^* \circ (I{\cdot}D)[g;\sigma;\omega] \circ \kappa_* \tag{2-36}$$

for any diffeomorphism κ of X. Here $\kappa_* \overset{\mathrm{def}}{=} (\kappa^{-1})^*$.

Proof. The first claim follows by a straightforward computation involving the standard identities

$$\begin{aligned} \Delta_{e^{2\varphi}g} &= e^{-2\varphi}\big[\Delta_g + (n-1)g(\mathrm{grad}_g(\varphi), \mathrm{grad}_g(\cdot))\big],\\ \mathsf{J}(e^{2\varphi}g) &= e^{-2\varphi}\Big[\mathsf{J}(g) - \Delta_g\varphi - \frac{n-1}{2}|d\varphi|_g^2\Big],\\ \Delta_g(e^{\omega\varphi}) &= \omega e^{\omega\varphi}\Delta_g\varphi + \omega^2 e^{\omega\varphi}|d\varphi|_g^2,\\ \mathrm{grad}_{e^{2\varphi}g}(f) &= e^{-2\varphi}\mathrm{grad}_g(f),\\ \Delta_g(f_1f_2) &= \Delta_g(f_1)f_2 + f_1\Delta_g(f_2) + 2g(\mathrm{grad}_g(f_1), \mathrm{grad}_g(f_2)),\\ \mathrm{grad}_g(f_1f_2) &= \mathrm{grad}_g(f_1)f_2 + f_1\mathrm{grad}_g(f_2) \end{aligned} \tag{2-37}$$

for any $f, f_1, f_2 \in C^\infty(X)$. These relations imply

$$
\begin{aligned}
&e^{-(\omega-1)\varphi} \circ (I{\cdot}D)[e^{2\varphi}g, e^{\varphi}\sigma; \omega] \circ e^{\omega\varphi} \\
&= (I{\cdot}D)[g,\sigma;\omega] + \Big[-\omega - \frac{\omega}{n+1}(2\omega+n-1) + \frac{2\omega}{n+1}(n+\omega)\Big]\sigma\Delta_g\varphi \\
&\qquad + \big[-(2\omega+n-1) + (2\omega+n-1)\big]\sigma g(\mathrm{grad}_g(\varphi), \mathrm{grad}_g(\cdot)) \\
&\quad + \Big[-\omega(\omega+n-1) + \omega(2\omega+n-1) - \frac{n\omega(2\omega+n-1)}{n+1} + \frac{\omega(n-1)(n+\omega)}{n+1}\Big]\sigma|d\varphi|_g^2 \\
&\qquad + \Big[\omega(2\omega+n-1) - (n+1)\frac{\omega(2\omega+n-1)}{n+1}\Big] g(\mathrm{grad}_g(\varphi), \mathrm{grad}_g(\sigma)) \\
&= (I{\cdot}D)[g,\sigma;\omega].
\end{aligned}
$$

The second claim is immediate by construction. The proof is complete. □

For (X, g) being the flat Euclidean space $(\mathbb{R}^{n+1}, g_0)$ and the scale r with zero-locus $\mathbb{R}^n$ (as in Section 2E), the degenerate Laplacian $I{\cdot}D$ reduces to

$$(I{\cdot}D)[g_0, r; \lambda-n-1] = -r\Delta_{g_0} + (2\lambda-n-3)\partial_r.$$

Hence

$$P(\lambda) = -(I{\cdot}D)[g_0; r, \lambda-n-1], \tag{2-38}$$

i.e., the shift operator $P(\lambda)$ is a special case of $I{\cdot}D$.

3. A curved version of the shift operator

Now we extend the definition of the shift operators $P(\lambda)$ (in Section 2E) to the setting of Section 2C. Thus, we assume that (M^n, h) is a Riemannian manifold of dimension $n \geq 3$ and we let $g_+ = r^{-2}(dr^2 + h_r)$ be an even Poincaré metric in normal form relative to h on $M_+^\circ = (0, \varepsilon) \times M$ (for some $\varepsilon > 0$). In particular, that means that, for odd n, h_r has an expansion

$$h_r = h + r^2 h_2 + r^4 h_4 + \cdots,$$

where all coefficients h_{2j} are determined by h, and for even n in the expansion

$$h_r = h + r^2 h_2 + \cdots + r^n h_n + r^{n+2} h_{n+2} + \cdots$$

the coefficients $h_{\leq n-2}$ and $\mathrm{tr}_h(h_n)$ are determined by h. We recall the notation $\bar{g} = r^2 g_+$. The following definition corresponds to (1-4).

Definition 3.1. The second-order differential operator

$$S(g_+; \lambda) \stackrel{\text{def}}{=} r\Delta_{\bar{g}} - (2\lambda-n+1)\partial_r - \tfrac{1}{2}(\lambda-n+1)\,\mathrm{tr}(h_r^{-1}\dot{h}_r) \tag{3-1}$$

is called a *shift operator*. Here the dot denotes the derivative with respect to r.

The notion *shift operator* is motivated by Theorem 3.7.

We shall regard $S(g_+;\lambda)$ as an operator $C^\infty(M_+^\circ) \to C^\infty(M_+^\circ)$. Since $\bar{g}$ is a smooth metric on $M_+ = [0,\varepsilon) \times M$, the shift operators may also be regarded as operators on $C^\infty(M_+)$, and mapping properties on M_+° naturally extend to M_+.

We emphasize that the shift operator $S(g_+;\lambda)$ is not completely determined by the Taylor coefficients of h_r. The composition $\iota^* S(g_+;\lambda)$ degenerates to a first-order operator.

By (2-16), the shift operator can also be written in the form

$$S(g_+;\lambda) \stackrel{\text{def}}{=} r\Delta_{\bar{g}} - (2\lambda-n+1)\partial_r - (\lambda-n+1)\dot{v}(r)/v(r). \tag{3-2}$$

The following result describes the relations among the three operators $\mathcal{D}_\lambda$, $(I\cdot D)[\cdot;\lambda]$ and $S(\cdot;\lambda)$.

Proposition 3.2. *It holds that*

$$S(g_+;\lambda) = -\mathcal{D}_{\lambda+1}(g_+). \tag{3-3}$$

Moreover, we have

$$S(g_+;\lambda) = -(I\cdot D)[\bar{g}; r, \lambda-n+1] \tag{3-4}$$

if g_+ is Einstein.

Proof. We recall the expressions (2-18) and (2-32) for Δ_{g_+} and $\mathcal{D}_\lambda(g_+)$. Moreover, by (2-19), we have

$$\Delta_{\bar{g}} = r^{-2}(\Delta_{g_+} + (n-1)r\partial_r).$$

Combining these formulas, we obtain

$$\begin{aligned} S(g_+;\lambda) &= r^{-1}(\Delta_{g_+} + (n-1)r\partial_r) - (2\lambda-n+1)\partial_r - \tfrac{1}{2}(\lambda-n+1)\operatorname{tr}(h_r^{-1}\dot{h}_r) \\ &= r\Delta_{h_r} + r\partial_r^2 + \tfrac{1}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r)r\partial_r - (2\lambda-n+1)\partial_r - \tfrac{1}{2}(\lambda-n+1)\operatorname{tr}(h_r^{-1}\dot{h}_r) \\ &= r\Delta_{h_r} + r\partial_r^2 - \Big(2\lambda-n+1 - \frac{r}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r)\Big)\partial_r - \tfrac{1}{2}(\lambda-n+1)\operatorname{tr}(h_r^{-1}\dot{h}_r) \\ &= -\mathcal{D}_{\lambda+1}(g_+). \end{aligned}$$

This proves the first claim. The second claim follows by evaluating $I\cdot D$ for $g = \bar{g}$ and $\sigma = r$. In particular, using

$$\Delta_{\bar{g}}(r) = \tfrac{1}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r)$$

and the relation

$$\mathsf{J}(\bar{g}) = -\frac{1}{2r}\operatorname{tr}(h_r^{-1}\dot{h}_r) \tag{3-5}$$

[Juhl 2009, (6.11.8)], we find

$$(I\cdot D)[\bar{g}; r, \omega] = -r\Delta_{\bar{g}} + (n+2\omega-1)\Big(\partial_r - \frac{\omega}{n+1}\Delta_{\bar{g}}(r)\Big) - \frac{2\omega}{n+1}(n+\omega)r\,\mathsf{J}(\bar{g})$$
$$= -r\Delta_{\bar{g}} + (n+2\omega-1)\partial_r + \frac{\omega}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r).$$

Hence $S(g_+;\lambda) = -(I\cdot D)[\bar{g}; r, \lambda-n+1]$. The proof is complete. □

Remark 3.3. The identity (3-5) is not valid for general Poincaré metrics g_+ since these are only asymptotically Einstein. For the convenience of the reader, we insert a proof of (3-5) for Einstein g_+ which also clarifies the modification in the general case. We recall the well-known transformation formula

$$\tau(\hat{g}) = e^{-2\varphi}\tau(g) - 2n\Delta_{\hat{g}}(\varphi) + n(n-1)|d\varphi|^2_{\hat{g}}, \quad \hat{g} = e^{2\varphi}g \tag{3-6}$$

for the scalar curvature. We apply (3-6) to $\hat{g} = dr^2+h_r$, $g = r^{-2}(dr^2+h_r)$ and $\varphi = \log r$. For Einstein g, i.e., $\operatorname{Ric}(g) + ng = 0$, we have $\tau(g) = -n(n+1)$. Hence we find

$$\tau(dr^2+h_r) = -\frac{n(n+1)}{r^2} - 2n\Delta_{dr^2+h_r}(\log r) + n(n-1)|d\log r|^2_{dr^2+h_r}$$
$$= -\frac{n(n+1)}{r^2} - 2n\Big(\partial_r^2(\log r) + \tfrac{1}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r)\partial_r(\log r)\Big) + \frac{n(n-1)}{r^2}$$
$$= -n\operatorname{tr}(h_r^{-1}\dot{h}_r)\frac{1}{r}.$$

Hence

$$\mathsf{J}(dr^2+h_r) = -\frac{1}{2r}\operatorname{tr}(h_r^{-1}\dot{h}_r).$$

We also note that, in [Gover and Waldron 2014], the authors deal with Einstein metrics outside hypersurfaces in Riemannian manifolds. In particular, the calculation at the end of [Gover and Waldron 2014, Section 5] employs the formula (3-5).

Remark 3.4. By the first part of Proposition 3.2, we have

$$S(g_+;\lambda) = r^{\lambda-n}\circ(\Delta_{g_+} + (\lambda+1)(n-\lambda-1))\circ r^{n-\lambda-1} \tag{3-7}$$

as an identity of operators acting on $C^\infty(M_+^\circ)$. Sometimes, this formula for $S(g_+;\lambda)$ will be more convenient to work with than the original definition (3-1). However, the original definition has the advantage that it clearly shows that $S(g_+;\lambda)$ acts on smooth functions on M_+. In order to illustrate the convenience of the conjugation formula (3-7), we note that it yields a slightly more conceptual proof of (3-5). We use the fact that for Einstein g_+,

$$P_2(g_+) = \Delta_{g_+} + m(m-1)$$

with $m = \frac{n+1}{2}$. Hence

$$P_2(\bar{g}) = r^{-m-1}P_2(g_+)r^{m-1} = r^{-m-1}(\Delta_{g_+} + m(m-1))r^{m-1} = r^{-1}S\Big(g_+;\frac{n-1}{2}\Big).$$

This identity is the special case $N=1$ of Theorem 3. On the function 1 it yields (3-5).

Remark 3.5. For even n, the condition $\mathrm{Ric}(g_+)+ng_+ = O(r^{n-2})$ implies $\tau(g_+) = -n(n+1)+O(r^n)$. Hence an extension of the arguments in Remark 3.4 gives

$$P_2(g_+) = \Delta_{g_+} + m(m-1) + O(r^n) \quad \text{and} \quad P_2(\bar{g}) = r^{-1} S\Big(g_+; \frac{n-1}{2}\Big) + O(r^{n-2}).$$

The trace condition (2-11) improves the estimate of the scalar curvature to $\tau(g_+) = -n(n+1)+o(r^n)$ and we obtain

$$r P_2(\bar{g}) = S\Big(g_+; \frac{n-1}{2}\Big) + o(r^{n-1}).$$

This is the special case $N=1$ of Theorem 3 for Poincaré metrics.

We continue with the discussion of the *shift property* of the shift operators. Its description requires some more notation. We recall that $M_+^\circ = (0,\varepsilon) \times M$ and $u \in C^\infty(M_+^\circ)$ be a solution to the equation

$$\Delta_{g_+} u + \nu(n-\nu)u = 0 \quad \text{for } \nu \in n/2 + i\mathbb{R},\ \nu \neq n/2 \tag{3-8}$$

with *boundary value* $f \in C^\infty(M)$ (see Section 2C). Such an *exact* eigenfunction can be constructed as described in Section 2C by regarding M_+ as part of a conformally compact manifold. Then u gives rise to the family (see [Juhl 2009, Section 6.6])

$$\lambda \mapsto M_u(r;\lambda) \in C^\infty(M_+^\circ), \qquad M_u(r;\lambda) \stackrel{\text{def}}{=} r^{\lambda-n+1} u, \quad \lambda \in \mathbb{C}. \tag{3-9}$$

We shall consider $M_u(r;\lambda)$ as a family of functions on M_+° as well as a family of distributions on M_+. In the latter case, we have

$$\langle M_u(r;\lambda), \varphi \rangle = \int_{M_+} r^{\lambda-n+1} u\varphi \, d\mathrm{vol}(\bar{g}), \quad \Re(\lambda) > n/2-2,$$

where the test functions φ are in $C_c^\infty(M_+)$ and the condition $\Re(\lambda) > \frac{n}{2}-2$ guarantees the convergence of the integral. Then $M_u(r;\lambda)$ will be understood as a meromorphic family of distributions. For simplicity, we shall denote that family of distributions also by $M_u(r;\lambda)$. Finally, we introduce the multiplication operators

$$M_r \stackrel{\text{def}}{=} r\cdot\,. \tag{3-10}$$

Remark 3.6. We consider the distribution $M_u(r;\lambda)$ in the flat case. Let $u \in C^\infty(\mathbb{H}^{n+1})$ be an eigenfunction of the Laplacian on the hyperbolic upper-half space $\mathbb{H}^{n+1}$. We assume that u can be written as Helgason's Poisson transform [1984],

$$u(r,x) = \int_{\mathbb{R}^n} \left(\frac{r}{|x-y|^2+r^2}\right)^\nu f(y)\,dy \tag{3-11}$$

of $f \in C_0^\infty(\mathbb{R}^n)$, say. Then

$$\begin{aligned}\langle M_u(r;\lambda),\varphi\rangle &= \int_{\mathbb{H}^{n+1}} r^{\lambda-n+1} u(r,x)\varphi(r,x)\,dr\,dx\\ &= \int_{\mathbb{H}^{n+1}}\int_{\mathbb{R}^n} r^{\lambda-n+1}\left(\frac{r}{|x-y|^2+r^2}\right)^{\nu} f(y)\varphi(r,x)\,dy\,dr\,dx\\ &= \int_{\mathbb{R}^n} f(y)\left(\int_{\mathbb{H}^{n+1}} K^+_{\lambda-2,\nu}(r,x-y)\varphi(r,x)\,dx\,dr\right)dy.\end{aligned}$$

The latter formula shows in which sense (3-9) can be regarded as a generalization of the distributional kernels $K^{\pm}_{\lambda,\nu}(r,x)$ (see (2-29)). We also note that the asymptotic expansion of u given by (3-11) is of the form

$$r^\nu \sum_{j\geq 0} r^{2j} a_{2j}(x) + r^{n-\nu} c(\lambda) \sum_{j\geq 0} r^{2j} b_{2j}(x), \quad r\to 0,$$

where $b_0 = f$ and $c(\lambda)$ is Harish-Chandra's c-function. In particular, the boundary value f appears in the coefficient of $r^{n-\lambda}$. Note that this notion of boundary value slightly differs from that used in Section 2C by the coefficient $c(\lambda)$.

Now we are ready to state and prove the *shift property* of the operators $S(g_+;\lambda)$. The following result generalizes [Fischmann et al. 2019, Theorem 3.5]. Following the terminology of [Fischmann et al. 2019], it may be referred to as a Bernstein–Sato identity.

Theorem 3.7. *Let (M^n, h) be a Riemannian manifold. Let u be a solution of* (3-8). *Then the multiplication operator M_r and the shift operator $S(g_+;\lambda)$ shift the λ-parameter when acting on the function $M_u(r;\lambda)$, i.e.*,

$$\begin{aligned}M_r(M_u(r;\lambda)) &= M_u(r;\lambda+1),\\ S(g_+;\lambda)(M_u(r;\lambda)) &= (\lambda+\nu-n+1)(\nu-\lambda-1)M_u(r;\lambda-1).\end{aligned}$$

Proof. The first claim is obvious from the definitions. Using (3-7) and the fact that u satisfies (3-8), we compute

$$\begin{aligned}S(g_+;\lambda)(M_u(r;\lambda)) &= r^{\lambda-n}\Delta_{g_+}(r^{n-\lambda-1}r^{\lambda-n+1}u) + (\lambda+1)(n-\lambda-1)r^{\lambda-n}u\\ &= -\nu(n-\nu)r^{\lambda-n}u + (\lambda-n+1)(-\lambda-1)r^{\lambda-n}u\\ &= (\lambda+\nu-n+1)(\nu-\lambda-1)M_u(r;\lambda-1).\end{aligned}$$

The proof is complete. □

From Theorem 3.7 we easily deduce the poles of the meromorphic continuation of the holomorphic family of distributions

$$C_c^\infty(M_+) \ni \varphi \mapsto \int_{M_+} M_u(r;\lambda)\varphi\,d\mathrm{vol}(\bar{g}), \quad \Re(\lambda) > \frac{n}{2} - 2.$$

Corollary 3.8. *The family* $\lambda \mapsto M_u(r;\lambda)$ *is meromorphic with generically simple poles at*[7]

$$\lambda = -\nu+n-2-N \quad \textit{and} \quad \lambda = \nu-2-N \tag{3-12}$$

for $N \in \mathbb{N}_0$.

Proof. Indeed, Theorem 3.7 implies

$$M_u(r;\lambda) = \frac{1}{(\lambda+\nu-n+2)(\nu-\lambda-2)} S(g_+;\lambda+1)(M_u(r;\lambda+1)).$$

The distribution on the left-hand side of this identity is holomorphic in the half-plane $\Re(\lambda) > \frac{n}{2}-2$. The right-hand side provides a meromorphic continuation to $\Re(\lambda) > \frac{n}{2}-3$ with simple poles in $\lambda = -\nu+n+2$ and $\lambda = \nu-2$. The assertion follows by a repeated application of that argument. □

Alternatively, Corollary 3.8 can be proved by directly inserting the asymptotic expansion of u into the integral which defines the distribution $M_u(r;\lambda)$ for $\Re(\lambda) > \frac{n}{2}-2$. The above argument using Theorem 3.7 is more conceptual, however.

Next, we discuss a conformal transformation law for shift operators. Let $\hat{h} = e^{2\varphi}h$ be conformally equivalent to h. We choose an even Poincaré metric g_+ in normal form relative to h and let $\hat{g}_+ = \kappa^*(g_+)$ be a related even Poincaré metric in normal form relative to $\hat{h}$; for the construction of the diffeomorphism κ we refer to Section 2C. In these terms, we have the following result.

Proposition 3.9. $S(\hat{g}_+;\lambda) = \left(\frac{r}{\kappa^*(r)}\right)^{\lambda-n} \circ \kappa^* \circ S(g_+;\lambda) \circ \kappa_* \circ \left(\frac{r}{\kappa^*(r)}\right)^{n-\lambda-1}.$

Proof. By (3-7), we have

$$S(\hat{g}_+;\lambda) = r^{\lambda-n}\Delta_{\hat{g}_+} r^{n-\lambda-1} + (\lambda+1)(n-\lambda-1)\frac{1}{r}.$$

But $\kappa^*(g_+) = \hat{g}_+$ implies $\Delta_{\hat{g}_+} = \kappa^*\Delta_{g_+}\kappa_*$. Hence

$$\begin{aligned} S(\hat{g}_+;\lambda) &= r^{\lambda-n}\kappa^* r^{-\lambda+n} r^{\lambda-n}\Delta_{g_+} r^{n-\lambda-1} r^{\lambda-n+1}\kappa_*(r^{n-\lambda-1}\cdot) + (\lambda+1)(n-\lambda-1)\frac{1}{r} \\ &= r^{\lambda-n}\kappa^*(r^{-\lambda+n})\kappa^* r^{\lambda-n}\Delta_{g_+} r^{n-\lambda-1}\kappa_*\kappa^*(r^{\lambda-n+1}) r^{n-\lambda-1} + (\lambda+1)(n-\lambda-1)\frac{1}{r} \\ &= \left(\frac{r}{\kappa^*(r)}\right)^{\lambda-n}\kappa^*(r^{\lambda-n}\Delta_{g_+} r^{n-\lambda-1})\kappa_*\left(\frac{r}{\kappa^*(r)}\right)^{n-\lambda-1} + (\lambda+1)(n-\lambda-1)\frac{1}{r}. \end{aligned}$$

Now the obvious identity

$$\frac{1}{r} = \left(\frac{r}{\kappa^*(r)}\right)^{\lambda-n} \circ \kappa^* \circ \frac{1}{r} \circ \kappa_* \circ \left(\frac{r}{\kappa^*(r)}\right)^{n-\lambda-1}$$

[7]Here ν is generic if both ladders of poles in (3-12) do not intersect. For $\nu \in \frac{n}{2}+i\mathbb{R}$ being *generic* it suffices to assume that $\nu \neq \frac{n}{2}$.

implies

$$S(\hat{g}_+, r; \lambda) = \Big(\frac{r}{\kappa^*(r)}\Big)^{\lambda-n} \kappa^* S(g_+; \lambda)\kappa_* \Big(\frac{r}{\kappa^*(r)}\Big)^{n-\lambda-1}.$$

This completes the proof. □

Theorem 1 is a direct consequence of this result.

Remark 3.10. For Einstein g_+, Proposition 3.9 also follows by combining the conformal covariance of $I{\cdot}D$ (Proposition 2.2) with the identification of $S(g_+; \lambda)$ as a degenerate Laplacian (Proposition 3.2). Indeed, for $\hat{h} = e^{2\varphi}h$, we write $g_+ = r^{-2}(dr^2 + h_r)$ and $\hat{g}_+ = r^{-2}(dr^2 + \hat{h}_r)$ with $\hat{g}_+ = \kappa^*(g_+)$ and $h_0 = h,\ \hat{h}_0 = \hat{h}$. Then

$$\kappa^*(dr^2 + h_r) = \Big(\frac{\kappa^*(r)}{r}\Big)^2 (dr^2 + \hat{h}_r). \tag{3-13}$$

Now assume that g_+ is Einstein. Then

$$\begin{aligned}
&S(\hat{g}_+; \lambda)\\
&\quad = -(I{\cdot}D)[dr^2 + \hat{h}_r, r; \lambda-n+1] && \text{(by (3-4))}\\
&\quad = -(I{\cdot}D)\Big[\Big(\frac{\kappa^*(r)}{r}\Big)^{-2}\kappa^*(dr^2+h_r), \Big(\frac{\kappa^*(r)}{r}\Big)^{-1}\kappa^*(r); \lambda-n+1\Big] && \text{(by (3-13))}\\
&\quad = -\Big(\frac{r}{\kappa^*(r)}\Big)^{\lambda-n}(I{\cdot}D)\big[\kappa^*(dr^2+h_r), \kappa^*(r); \lambda-n+1\big]\Big(\frac{r}{\kappa^*(r)}\Big)^{\lambda-n+1} && \text{(by (2-35))}\\
&\quad = -\Big(\frac{r}{\kappa^*(r)}\Big)^{\lambda-n}\kappa^*(I{\cdot}D)\big[dr^2+h_r, r; \lambda-n+1\big]\kappa_*\Big(\frac{r}{\kappa^*(r)}\Big)^{\lambda-n+1} && \text{(by (2-36))}\\
&\quad = \Big(\frac{r}{\kappa^*(r)}\Big)^{\lambda-n}\kappa^* S(g_+; \lambda)\kappa_*\Big(\frac{r}{\kappa^*(r)}\Big)^{\lambda-n+1}. && \text{(by (3-4))}
\end{aligned}$$

A general Poincaré metric g_+ is only approximate Einstein. For such a metric, the identification of $S(g_+; \lambda)$ with a degenerate Laplacian holds true only up to an error term. This leads to error terms in the above calculation and in the resulting conformal transformation law.

In Section 4, we shall prove a relation between compositions of shift operators and residue families. For that purpose, we need a relation among the operator $S(g_+; \lambda)$ and its formal adjoint $S^*(g_+; \lambda)$ with respect to the scalar product defined by $\bar{g}$.

Proposition 3.11. *The formal adjoint of the operator $S(g_+; \lambda)$ acting on $C_c^\infty(M_+^\circ)$ with respect to the scalar product defined by $\bar{g}$ is given by*

$$S^*(g_+; \lambda) = S(g_+; n-\lambda-2).$$

Proof. Let $f_1, f_2 \in C_c^\infty(M_+^\circ)$. Then, using $d\mathrm{vol}(\bar{g}) = r^{n+1}\, d\mathrm{vol}(g_+)$ and (3-7), we have

$$\begin{aligned}\int_{M_+} (S(g_+;\lambda)f_1)f_2\, d\mathrm{vol}(\bar{g}) &= \int_{M_+} r^{\lambda-n}[\Delta_{g_+}+(\lambda+1)(n-\lambda-1)](r^{n-\lambda-1}f_1)f_2 r^{n+1}\, d\mathrm{vol}(g_+)\\ &= \int_{M_+} [\Delta_{g_+}+(\lambda+1)(n-\lambda-1)](r^{n-\lambda-1}f_1)r^{\lambda+1}f_2\, d\mathrm{vol}(g_+)\\ &= \int_{M_+} (r^{n-\lambda-1}f_1)[\Delta_{g_+}+(\lambda+1)(n-\lambda-1)](r^{\lambda+1}f_2)\, d\mathrm{vol}(g_+)\\ &= \int_{M_+} f_1(r^{-\lambda-2}[\Delta_{g_+}+(\lambda+1)(n-\lambda-1)](r^{\lambda+1}f_2))\, d\mathrm{vol}(\bar{g}).\end{aligned}$$

Thus

$$S^*(g_+;,\lambda) = r^{-\lambda-2}\circ(\Delta_{g_+}+(\lambda+1)(n-\lambda-1))\circ r^{\lambda+1}.$$

Comparing this relation with (3-7) completes the proof. □

In the proof of Theorem 4.1, we shall actually need the following improved version of the latter result.

Remark 3.12. The relation

$$\int_{M_+} S(g_+;\lambda)(f_1)f_2\, d\mathrm{vol}(\bar{g}) = \int_{M_+} f_1 S(g_+; n-\lambda-2)(f_2)\, d\mathrm{vol}(\bar{g})$$

continues to be true for $f_2 \in C_c^\infty(M_+)$ and $f_1 \in C^\infty(M_+^\circ)$ with an asymptotic expansion of the form

$$f_1(r,x) \sim \sum_{j\geq 0} r^{\nu+j} a_j(x), \quad r\to 0, \quad a_j \in C^\infty(M)$$

with $\Re(\nu) > 0$.

Proof. It suffices to prove the relations

$$\int_{M_+} \partial_r(f_1)f_2\, d\mathrm{vol}(\bar{g}) = -\int_{M_+} f_1\partial_r(f_2)\, d\mathrm{vol}(\bar{g}) - \tfrac{1}{2}\int_{M_+} f_1\,\mathrm{tr}(h_r^{-1}\dot{h}_r)f_2\, d\mathrm{vol}(\bar{g})$$

and

$$\int_{M_+} (r\Delta_{\bar{g}})(f_1)f_2\, d\mathrm{vol}(\bar{g}) = \int_{M_+} f_1\big(r\Delta_{\bar{g}} + 2\partial_r + \tfrac{1}{2}\mathrm{tr}(h_r^{-1}\dot{h}_r)\big)(f_2)\, d\mathrm{vol}(\bar{g}).$$

For the proof of the first relation we use (2-16) and the observation that the restriction of f_1 to the boundary vanishes. Green's formula and (2-19) imply the second relation. Here the boundary contributions vanish since the restrictions of f_1 and $r\partial_r(f_1)$ to the boundary both vanish. □

For later purpose, we need to understand the behavior of the composition of $S(g_+;\lambda)$ with the multiplication by powers of r.

Lemma 3.13. *Let $f \in C^\infty(M_+^\circ)$ and $a \in \mathbb{N}$. Then*

$$S(g_+;\lambda)(r^a f) = r^a S(g_+;\lambda{-}a)f - a(2\lambda{-}n{+}2{-}a)r^{a-1}f.$$

Proof. The proof is straightforward. The definition (3-1) and the formula (2-19) for $\Delta_{\bar{g}}$ imply

$$\begin{aligned} S(g_+;\lambda)(r^a f) &= r\Delta_{\bar{g}}(r^a f) - (2\lambda{-}n{+}1)\partial_r(r^a f) - \tfrac{1}{2}(\lambda{-}n{+}1)\operatorname{tr}(h_r^{-1}\dot{h}_r)r^a f \\ &= r^a S(g_+;\lambda{-}a)f + [a(a{-}1) - (2\lambda{-}n{+}1)]r^{a-1}f \\ &= r^a S(g_+;\lambda{-}a)f - a(2\lambda{-}n{+}2{-}a)r^{a-1}f. \end{aligned}$$

The proof is complete. □

Of course, this result is also true for $f \in C^\infty(M_+)$. In view of the second part of Proposition 3.2, it should be viewed as an analog of the $\mathfrak{sl}(2)$-structure for the degenerate Laplacian $I{\cdot}D$ proved in [Gover and Waldron 2014, Lemma 3.1, Proposition 3.4].

4. A new formula for residue families

We recall that residue families $D_N^{\text{res}}(h;\lambda)$ are defined by normalizations of the families $\delta_N(h;\lambda)$ (see (2-27)). In the present section, we express these families in terms of shift operators $S(g_+;\lambda)$ (see (3-1)) and discuss some consequences.

For $N \in \mathbb{N}$, we define the family $S_N(g_+;\lambda)$ of shift operators on M_+° by

$$S_N(g_+;\lambda) \stackrel{\text{def}}{=} \underbrace{S(g_+;\lambda)\circ\cdots\circ S(g_+;\lambda+N-1)}_{N \text{ factors}}. \tag{4-1}$$

We also set $S_0(g_+;\lambda) = \mathrm{Id}$.

Similarly, we define

$$(I{\cdot}D)_N[\bar{g},r;\lambda] \stackrel{\text{def}}{=} \underbrace{(I{\cdot}D)[\bar{g},r;\lambda]\circ\cdots\circ(I{\cdot}D)[\bar{g},r;\lambda+N-1]}_{N \text{ factors}}. \tag{4-2}$$

Again, we shall regard $S_N(g_+;\lambda)$ as an operator $C^\infty(M_+^\circ)\to C^\infty(M_+^\circ)$ and also as an operator on $C^\infty(M_+)$. By definition, $S_N(g_+;\lambda)$ is a differential operator of order $2N$ with polynomial coefficients in λ. As a polynomial in λ it is of degree N.

We also recall that, for $N \in \mathbb{N}$ with $N \le n+1$ for even n, the families $\delta_N(h;\lambda)$ are determined by h.

Theorem 4.1. *Let $N \in \mathbb{N}$ with $N \le n+1$ for even n. Then*

$$\delta_N(h;\lambda) = \frac{1}{(-N)_N(2\lambda{-}n{+}1)_N}\iota^* S_N(g_+;\lambda) \tag{4-3}$$

as an identity of meromorphic functions in λ *with values in operators* $C^\infty(M_+) \to C^\infty(M)$.

Proof. Let u be an eigenfunction with boundary value f and satisfying (3-8) with $\Re(\nu) = n/2$, $\nu \neq n/2$. We derive the assertion from Theorem 3.7 and the identity

$$\operatorname{Res}_{\lambda=-\nu-1}\left(\int_{M_+} M_u(r;\lambda+n-1)\varphi\, d\mathrm{vol}(\bar g)\right) = \int_M f\iota^*\varphi\, d\mathrm{vol}(h) \tag{4-4}$$

(see (2-27) for $N = 0$). In the following, it will be convenient to use, for any λ-dependent operator $A(\lambda)$, the notation

$$A((\lambda)_N) \stackrel{\text{def}}{=} A(\lambda)\circ A(\lambda+1)\circ\cdots\circ A(\lambda+N-1),$$
$$A((\lambda)^N) \stackrel{\text{def}}{=} A(\lambda)\circ A(\lambda-1)\circ\cdots\circ A(\lambda-N+1).$$

On the one hand, (2-27) implies

$$\operatorname{Res}_{\lambda=-\nu-1-N}\left(\int_{M_+} M_u(r;\lambda+n-1)\varphi\, d\mathrm{vol}(\bar g)\right) = \int_M f\delta_N(h;\nu)\varphi\, d\mathrm{vol}(h). \tag{4-5}$$

On the other hand, using Theorem 3.7, we obtain for $\Re(\lambda) \notin -\frac{n}{2} - \mathbb{N}$

$$\begin{aligned} M_u(r;\lambda+n-1) &= \frac{S(g_+;\lambda+n)(M_u(r;\lambda+n))}{(\lambda+\nu+1)(\nu-\lambda-n-1)}\\ &= \frac{S(g_+;\lambda+n)S(g_+;\lambda+n+1)(M_u(r;\lambda+n+1))}{(\lambda+\nu+1)(\lambda+\nu+2)(\nu-\lambda-n-2)(\nu-\lambda-n-1)}\\ &= \cdots = \frac{S(g_+;(\lambda+n)_N)(M_u(r;\lambda+n+N-1))}{(\lambda+\nu+1)_N(\nu-\lambda-n-N)_N}. \end{aligned}$$

Now we take adjoints using Remark 3.12. This gives

$$\begin{aligned} \int_{M_+} M_u(r;\lambda+n-1)\varphi\, d\mathrm{vol}(\bar g) &= \frac{1}{(\lambda+\nu+1)_N(\nu-\lambda-n-N)_N}\\ &\quad\times\int_{M_+} M_u(r;\lambda+n+N-1)S^*(g_+;(\lambda+n+N-1)^N)\varphi\, d\mathrm{vol}(\bar g) \end{aligned}$$

for $\Re(\lambda) > -\frac{n}{2}-1$. By the assumptions, the zeros of $\nu \mapsto (\lambda+\nu+1)_N(\nu-\lambda-n-N)_N$ are simple for $\Re(\lambda) > -\frac{n}{2}-1$. Hence (4-4) implies

$$\begin{aligned} &\operatorname{Res}_{\lambda=-\nu-1-N}\left(\int_{M_+} M_u(r;\lambda+n-1)\varphi\, d\mathrm{vol}(\bar g)\right)\\ &\qquad= \frac{1}{(-N)_N(2\nu-n+1)_N}\int_M f\iota^*S^*(g_+;(-\nu+n-2)^N)\varphi\, d\mathrm{vol}(h)\\ &\qquad= \frac{1}{(-N)_N(2\nu-n+1)_N}\int_M f\iota^*S_N(g_+;\nu)\varphi\, d\mathrm{vol}(h). \end{aligned}$$

Comparing this result with (4-5), completes the proof for $\Re(\nu) = n/2,\ \nu \neq n/2$. The assertion then follows by meromorphic continuation. □

Since the operator $S_N(g_+; \lambda)$ involves $2N$ derivatives in r, it is a nontrivial observation that its composition with ι^* only depends on h for appropriate choices of N. But this is an immediate consequence of Theorem 4.1 as long as $N \leq n$ for even n. Therefore, it is justified to introduce the notation

$$\Sigma_N(h; \lambda) \stackrel{\text{def}}{=} \iota^* S_N(g_+; \lambda) \tag{4-6}$$

for such N. However, for even n and general $N \in \mathbb{N}$, the composition $\iota^* S_N(g_+; \lambda)$ does not only depend on h and the notation $\Sigma_N(h; \lambda)$ will not be used.

As a direct consequence of Theorem 4.1, we obtain the following identification of residue families with operators $\Sigma_N(h; \lambda)$.

Corollary 4.2. *Assume that $N \in \mathbb{N}$ so that $2N \leq n$ for even n. Then*

$$D_{2N}^{\text{res}}(h; \lambda) = \frac{1}{(-2N)_N \left(\lambda + \frac{n}{2} - 2N + \frac{1}{2}\right)_N} \Sigma_{2N}(h; \lambda + n - 2N) \tag{4-7}$$

and

$$D_{2N+1}^{\text{res}}(h; \lambda) = \frac{1}{2(-2N-1)_{N+1} \left(\lambda + \frac{n}{2} - 2N - \frac{1}{2}\right)_{N+1}} \Sigma_{2N+1}(h; \lambda + n - 2N - 1). \tag{4-8}$$

Some further comments on this result are in order.

We recall that all residue families $D_N^{\text{res}}(h; \lambda)$ are polynomials in λ. Hence Corollary 4.2 shows that the zeros of the denominators on the right-hand sides of (4-7) and (4-8) actually are zeros of the respective numerators. These formulas naturally reflect the degrees of the families on both sides. In fact, the degrees of the polynomials $D_{2N}^{\text{res}}(h; \lambda)$ and $\Sigma_{2N}(h; \lambda)$ are N and $2N$, respectively. Similarly, the degrees of $D_{2N+1}^{\text{res}}(h; \lambda)$ and $\Sigma_{2N+1}(h; \lambda)$ are N and $2N + 1$, respectively.

Corollary 4.2 also shows that the composition of the order $2N$ family $S_N(g_+; \lambda)$ with ι^* degenerates to a family of order N. Some of the properties of the families $S_N(g_+; \lambda)$ away from $r = 0$ will be discussed in Sections 5–6.

A direct consequence of Corollary 4.2 and Lemma 4.6 are factorization identities for residue families $D_N^{\text{res}}(h; \lambda)$ into compositions with the factors

$$S(g_+; \lambda + n - 1) \quad \text{and} \quad M_r.$$

Corollary 4.3. *Let $N \in \mathbb{N}$ so that $2N \leq n$ for even n. Then*

$$\begin{aligned} D_{2N}^{\text{res}}(h; \lambda) &= D_{2N-1}^{\text{res}}(h; \lambda - 1) S(g_+; \lambda + n - 1), \\ -(2N+1)(2\lambda + n - 2N - 1) D_{2N+1}^{\text{res}}(h; \lambda) &= D_{2N}^{\text{res}}(h; \lambda - 1) S(g_+; \lambda + n - 1), \end{aligned}$$

and
$$D^{\mathrm{res}}_{2N}(h;\lambda) = D^{\mathrm{res}}_{2N+1}(h;\lambda+1)M_r,$$
$$-2N(2\lambda+n-2N+2)D^{\mathrm{res}}_{2N-1}(h;\lambda) = D^{\mathrm{res}}_{2N}(h;\lambda+1)M_r.$$

Proof. The first two identities immediately follow from Corollary 4.2. The proofs of the last two identities also require Lemma 4.6. We omit the details. □

The following result is a consequence of the conformal transformation law in Proposition 3.9.

Lemma 4.4. *Let $N \in \mathbb{N}$ and assume that $\hat{h} = e^{2\varphi}h$. Then*
$$S_N(\hat{g}_+;\lambda) = \Big(\frac{\kappa^*(r)}{r}\Big)^{n-\lambda}\kappa^* S_N(g_+;\lambda)\kappa_*\Big(\frac{\kappa^*(r)}{r}\Big)^{\lambda+N-n}.$$

Proof. Proposition 3.9 yields
$$S(\hat{g}_+;\lambda) = \Big(\frac{\kappa^*(r)}{r}\Big)^{n-\lambda}\kappa^* S(g_+;\lambda)\kappa_*\Big(\frac{\kappa^*(r)}{r}\Big)^{\lambda-n+1}.$$
An application of that identity to the composition $S_N(\hat{g}_+;\lambda)$ gives
$$S_N(\hat{g}_+;\lambda) = \Big(\frac{\kappa^*(r)}{r}\Big)^{n-\lambda}\kappa^* S(g_+;\lambda)\cdots S(g_+;\lambda+N-1)\kappa_*\Big(\frac{\kappa^*(r)}{r}\Big)^{\lambda+N-n}.$$
The proof is complete. □

Now by combining Theorem 4.1 with Lemma 4.4, we obtain an alternative proof of the conformal transformation law of residue families [Juhl 2009, Theorem 6.6.3].

Corollary 4.5. *Let $N \in \mathbb{N}$ so that $N \le n+1$ for even n. Assume that $\hat{h} = e^{2\varphi}h$. Then*
$$D^{\mathrm{res}}_N(\hat{h};\lambda) = e^{(\lambda-N)\varphi}D^{\mathrm{res}}_N(h;\lambda)\kappa_*\Big(\frac{\kappa^*(r)}{r}\Big)^{\lambda}.$$

Proof. The assertion is a consequence of Theorem 4.1, Lemma 4.4, the limit formula [Juhl 2009, (6.6.16)]
$$\lim_{r\to 0}\Big(\frac{\kappa^*(r)}{r}\Big) = e^{-\varphi}$$
and the fact that κ acts as the identity on M. □

Finally, we prove:

Lemma 4.6. *Let $N \in \mathbb{N}$ and $f \in C^\infty(M_+^\circ)$. Then*

(4-9) $$S_N(g_+;\lambda)(rf) = rS_N(g_+;\lambda-1)(f) - N(2\lambda-n+N)S_{N-1}(g_+;\lambda)(f).$$

Proof. We iteratively apply Lemma 3.13 to compute

$$
\begin{aligned}
&S_N(g_+;\lambda)(rf)\\
&\quad = S(g_+;\lambda)\circ\cdots\circ S(g_+;\lambda+N-1)(rf)\\
&\quad = S_{N-1}(g_+;\lambda) r S(g_+;\lambda+N-2)(f) - (2\lambda-n+2N-1)S_{N-1}(g_+;\lambda)(f)\\
&\quad = \cdots = rS_N(g_+;\lambda-1)(f) - \left[N(2\lambda-n+2N+1) - \sum_{j=1}^{N} 2j\right] S_{N-1}(g_+;\lambda)(f)
\end{aligned}
$$

Now the identity $\sum_{j=1}^N 2j = N(N+1)$ completes the proof. □

The identity (4-9) obviously holds true also for all $f \in C^\infty(M_+)$.

5. Applications

In the present section, we further exploit the relation between families of shift operators and residue families. The flow of information will be in both directions, i.e., we use facts on families of shift operators to derive properties of residue families and also use properties of residue families to derive properties of families of shift operators.

It was shown in [Juhl 2009; 2013] that residue families $D^{\mathrm{res}}_{2N}(h;\lambda)$ satisfy two systems of factorization identities. In Section 5A we provide new proofs of these identities. They rest on the identification of two special values of the families of shift operator in terms of GJMS operators. These are given in Theorem 5.1 and Theorem 5.6, respectively. Theorem 5.1 will be further exploited in Section 6F. In Section 5B, we shall describe a compressed formulation of the recursive algorithm for the solution operators $\mathcal{T}_{2j}(h;\lambda)$ in terms of the families $S_N(g_+;\lambda)$. In the last section, we derive a new formula for all Q-curvatures (critical and subcritical ones) in even dimension in terms of shift operators.

5A. ***Shift operators and GJMS operators.*** Juhl [2009; 2013] proved that, for $N \in \mathbb{N}$ with $2N \le n$ for even n, the even-order residue families $D^{\mathrm{res}}_{2N}(h;\nu)$ satisfy the identities

$$
(5\text{-}1)\quad D^{\mathrm{res}}_{2N}\Big(h; -\frac{n}{2}+N\Big) = P_{2N}(h)\iota^* \quad\text{and}\quad D^{\mathrm{res}}_{2N}\Big(h; -\frac{n+1}{2}+N\Big) = \iota^* P_{2N}(\bar{g}).
$$

We briefly comment on the well-definedness of the second identity. For odd n, the Taylor coefficients of h_r are determined by h. Hence, for any $N \in \mathbb{N}$, the left-hand side of the second identity is determined by h. The right-hand side of this identity involves a GJMS operator in even dimension $n+1$. We recall that, for general metrics, these are only defined for subcritical orders $2N \le n+1$. But here they are defined for all $N \in \mathbb{N}$. In fact, there is an explicit formula for the Taylor coefficients of a Poincaré metric of $\bar{g}$ [Juhl 2013]. These are determined by h_r, i.e., by h. Thus the right-hand side is well-defined for all $N \in \mathbb{N}$. For even n, the left-hand side of the second identity is defined for $2N \leq n$. The right-hand side uses derivatives of a Poincaré metric for $\bar{g}$ which are determined by h.

For even N, the identities (5-1) are the simplest respective special cases of the systems

$$D_N^{\mathrm{res}}\Big(h; -\frac{n}{2}+N-k\Big) = P_{2k}(h) D_{N-2k}^{\mathrm{res}}\Big(h; -\frac{n}{2}+N-k\Big), \quad 2 \leq 2k \leq N \tag{5-2}$$

and

$$D_N^{\mathrm{res}}\Big(h; -\frac{n+1}{2}+k\Big) = D_{N-2k}^{\mathrm{res}}\Big(h; -\frac{n+1}{2}-k\Big) P_{2k}(\bar{g}), \quad 2 \leq 2k \leq N \tag{5-3}$$

of factorization identities [Juhl 2013, Theorems 3.1–3.2]. They play an important role in connection with the description of recursive structures among GJMS operators and Q-curvatures. Here we shall give an *independent proof* of the second system. The arguments will also prove their counterparts for odd-order residue families. Moreover, we derive the first system from its special case $k = N$, i.e., from the first identity in (5-1). The new proofs completely differ from earlier arguments.

We start with the proof of system (5-3). The proof rests on Corollary 4.2 and two basic facts. The first of these is also of independent interest. It will be used in Section 6F.

Theorem 5.1. *Let $N \in \mathbb{N}$ so that $2N \leq n$ if n is even. Set $m = \frac{n+1}{2}$. Then*

$$S_N(g_+; m-1) = r^N P_{2N}(\bar{g})$$

up to an error term in $O(r^\infty)$ for odd n and $o(r^{n-N})$ for even n. Moreover, the equality is true without an error term if g_+ is Einstein.

Proof. We first consider the case g_+ Einstein. The identity (2-4) shows that the $2N$-th order GJMS operator of g_+ is given by the product

$$P_{2N}(g_+) = \prod_{l=1}^{N} \big(\Delta_{g_+} + (m+l-1)(m-l)\big). \tag{5-4}$$

The conformal covariance of GJMS operators implies that

$$P_{2N}(\bar{g}) = r^{-m-N} P_{2N}(g_+) r^{m-N}.$$

By the definition (4-1), we have

$$S_N(g_+; m-1) = S(g_+; m-1) \circ \cdots \circ S(g_+; m+N-2).$$

Hence, using (3-7) and (5-4), we obtain

$$\begin{aligned} S_N(g_+; m-1) &= r^{-m}\big(\Delta_{g_+} + m(m-1)\big)\cdots\big(\Delta_{g_+} + (m+N-1)(m-N)\big)r^{m-N} \\ &= r^{-m} P_{2N}(g_+) r^{m-N} \\ &= r^N P_{2N}(\bar{g}). \end{aligned}$$

This completes the proof for g_+ Einstein. For general Poincaré metrics and odd n, the assertion follows by similar arguments using the generalization

$$P_{2N}(g_+) = \prod_{l=1}^{N}\big(\Delta_{g_+} + (m+l-1)(m-l)\big) + O(r^\infty) \tag{5-5}$$

of (5-4). The conformal covariance of P_{2N} show that

$$S_N(g_+; m-1) = r^N P_{2N}(\bar{g}) + O(r^\infty).$$

Similarly, for even n, the formula

$$P_{2N}(g_+) = \prod_{l=1}^{N}\big(\Delta_{g_+} + (m+l-1)(m-l)\big) + o(r^n) \tag{5-6}$$

(see Remark 3.5 for $N = 1$) and the conformal covariance of P_{2N} show that

$$S_N(g_+; m-1) = r^N P_{2N}(\bar{g}) + o(r^{n-N}).$$

The proof is complete. □

Theorem 5.1 directly implies the following N factorization identities.

Corollary 5.2. *Let $N \in \mathbb{N}$ so that $2N \le n$ for even n. Let $0 \le k \le N-1$. Then*

$$S_N(g_+; m-k-1) = S_k(g_+; m-k-1) r^{N-k} P_{2N-2k}(\bar{g})$$

if g_+ is Einstein. For general Poincaré metrics g_+, the identity holds true with an error term in $O(r^\infty)$ for odd n and

$$S_k(g_+; m-k-1) o(r^{n-N+k})$$

for even n.

The second basic fact is a generalization of Lemma 3.13 which states that

$$S(g_+; \lambda) r^j = r^j S(g_+; \lambda-j) - j(2\lambda-n-j+2) r^{j-1}$$

for $j \in \mathbb{N}$.

Lemma 5.3. *Let $k, j \in \mathbb{N}$. Then*

$$S_k(g_+; \lambda) r^j = \sum_{l=0}^{k} \binom{k}{l} (-j)_l (2\lambda-n-j+k+1)_l r^{j-l} S_{k-l}(g_+; \lambda-j+l). \tag{5-7}$$

Proof. We use induction over k. For $k = 1$, we have

$$S_1(g_+;\lambda)r^j = S(g_+;\lambda)r^j = r^j S(g_+;\lambda-j) - j(2\lambda-n-j+2)r^{j-1}$$
$$= \sum_{l=0}^{1}\binom{1}{l}(-j)_l(2\lambda-n-j+2)_l r^{j-l} S_{1-l}(g_+;\lambda-j+l).$$

Now assume that the assertion holds true for $k-1$. Then we compute

$$\begin{aligned}
&S_k(g_+;\lambda)r^j \\
&\quad = S(g_+;\lambda)\circ S_{k-1}(g_+;\lambda+1)r^j \\
&\quad = S(g_+;\lambda)\sum_{l=0}^{k-1}\binom{k-1}{l}(-j)_l(2\lambda-n-j+k+2)_l r^{j-l} S_{k-l-1}(g_+;\lambda-j+l+1) \\
&\quad = \sum_{l=0}^{k-1}\binom{k-1}{l}(-j)_l(2\lambda-n-j+k+2)_l S(g_+;\lambda) r^{j-l} S_{k-l-1}(g_+;\lambda-j+l+1).
\end{aligned}$$

By Lemma 3.13, the last display equals

$$\begin{aligned}
&\sum_{l=0}^{k-1}\binom{k-1}{l}(-j)_l(2\lambda-n-j+k+2)_l r^{j-l} S(g_+;\lambda-j+l) S_{k-l-1}(g_+;\lambda-j+l+1) \\
&\qquad - \sum_{l=0}^{k-1}\binom{k-1}{l}(-j)_l(2\lambda-n-j+k+2)_l \\
&\qquad\qquad \times (j-l)(2\lambda-n-j+l+2) r^{j-l-1} S_{k-l-1}(g_+;\lambda-j+l+1).
\end{aligned}$$

Shifting the summation index in the second sum and simplification gives

$$\begin{aligned}
&\sum_{l=0}^{k-1}\binom{k-1}{l}(-j)_l(2\lambda-n-j+k+2)_l r^{j-l} S_{k-l}(g_+;\lambda-j+l) \\
&\quad + \sum_{l=1}^{k}\binom{k-1}{l-1}(-j)_l(2\lambda-n-j+k+2)_{l-1}(2\lambda-n-j+l+1) r^{j-l} S_{k-l}(g_+;\lambda-j+l).
\end{aligned}$$

Now observe that

$$(2\lambda-n-j+k+2)_l = (2\lambda-n-j+k+1)_l + l(2\lambda-n-j+k+2)_{l-1}$$

and

$$\begin{aligned}
&(2\lambda-n-j+k+2)_{l-1}(2\lambda-n-j+l+1) \\
&\qquad = (2\lambda-n-j+k+1)_l - (k-l)(2\lambda-n-j+k+2)_{l-1}
\end{aligned}$$

as well as

$$l\binom{k-1}{l}-(k-l)\binom{k-1}{l-1}=0 \quad \text{and} \quad \binom{k-1}{l}-\binom{k-1}{l-1}=\binom{k}{l}.$$

Putting things together, we conclude

$$S_k(g_+;\lambda)r^j=\sum_{l=0}^{k}\binom{k}{l}(-j)_l(2\lambda-n-j+k+1)_l r^{j-l}S_{k-l}(g_+;\lambda-j+l).$$

The proof is complete. □

Now we are able to prove the system (5-3).

Theorem 5.4. *Let* $N\in\mathbb{N}$ *so that* $N\leq n+1$ *for even* n*. Then*

$$D_N^{\mathrm{res}}\Big(h;-\frac{n+1}{2}+k\Big)=D_{N-2k}^{\mathrm{res}}\Big(h;-\frac{n+1}{2}-k\Big)P_{2k}(\bar{g}) \tag{5-8}$$

for all $0\leq 2k\leq N$.

Proof. We only discuss the case of even N. The proof in the odd-order case is analogous. By Corollary 4.2, we have

$$D_{2N}^{\mathrm{res}}\Big(h;-\frac{n+1}{2}+k\Big)=\frac{1}{(-2N)_N(k-2N)_N}\iota^*S_{2N}\Big(g_+;\frac{n-1}{2}+k-2N\Big).$$

Now we additionally assume that g_+ is Einstein. Then Corollary 5.2 implies the identity

$$S_{2N}\Big(g_+;\frac{n-1}{2}+k-2N\Big)=S_{2N-k}\Big(g_+;\frac{n-1}{2}+k-2N\Big)r^kP_{2k}(\bar{g}) \tag{5-9}$$

We use Lemma 5.3 to conclude that

$$\begin{aligned}S_{2N-k}\Big(g_+;\frac{n-1}{2}+k-2N\Big)r^k&\\ &=\sum_{l=0}^{2N-k}\binom{2N-k}{l}(-k)_l(-2N)_l r^{k-l}S_{2N-k-l}\Big(g_+;\frac{n-1}{2}+l-2N\Big).\end{aligned}$$

Note that no summand with summation index $l\geq k+1$ will contribute due to $(-k)_l=0$. By restriction to $r=0$, the last display yields

$$\begin{aligned}\iota^*S_{2N-k}\Big(g_+;\frac{n-1}{2}+k-2N\Big)r^k&\\ &=\binom{2N-k}{k}(-k)_k(-2N)_k\iota^*S_{2N-2k}\Big(g_+;\frac{n-1}{2}+k-2N\Big).\end{aligned} \tag{5-10}$$

Finally, by Corollary 4.2, we have

$$D^{\text{res}}_{2N-2k}\Big(h; -\frac{n+1}{2}-k\Big) = \frac{1}{(-2N+2k)_{N-k}(k-2N)_{N-k}} \iota^* S_{2N-2k}\Big(g_+; \frac{n-1}{2}+k-2N\Big).$$

Combining these observation, proves

$$D^{\text{res}}_{2N}\Big(h; -\frac{n+1}{2}+k\Big) = \frac{\binom{2N-k}{k}(-k)_k(-2N)_k(-2N+2k)_{N-k}(k-2N)_{N-k}}{(-2N)_N(k-2N)_N} \times D^{\text{res}}_{2N-2k}\Big(h; -\frac{n+1}{2}-k\Big) P_{2k}(\bar{g}).$$

Now the combinatorial identity

$$\frac{\binom{2N-k}{k}(-k)_k(-2N)_k(-2N+2k)_{N-k}(k-2N)_{N-k}}{(-2N)_N(k-2N)_N} = 1$$

completes the proof for Einstein g_+. For general Poincaré metrics g_+, we have to control the error terms coming from Corollary 5.2. For odd n, error terms obviously do to not contribute to (5-10). For even n, the relation (5-9) contains an error term in

$$S_{2N-k}\Big(g_+; \frac{n-1}{2}+k-2N\Big) o(r^{n-k}).$$

By Lemma 5.3, this contribution is contained in $o(r^{n-2N})$. Hence its composition with ι^* vanishes if $2N \leq n$.[8] □

Remark 5.5. The proof of the factorizations (5-8) for even N given in [Juhl 2013] assumes that g_+ is Einstein. A closer inspection of this proof shows that it can be refined to establish the assertion in full generality. The refinement rests on the factorizations (5-5) and (5-6) with remainder terms. The point is that the remainder terms do not contribute to the residue calculations in the refinement of that proof. Note also that, along these lines, again only the critical case $2N = n$ (for n even) requires the remainder term $o(r^n)$ in (5-6).

We continue with the discussion of the factorization identities (5-2). Their proof rests on Corollary 4.2 and the identification of the value

$$\iota^* S_{2N}\Big(g_+; \frac{n}{2}-N\Big) = \Sigma_{2N}\Big(h, \frac{n}{2}-N\Big)$$

as a tangential operator being a multiple of $P_{2N}(h)\iota^*$. This fact actually will be deduced from the first identity in (5-1). Unfortunately, we do not have an

[8]The arguments show that, for even n and $2N < n$, the weaker estimate $O(r^n)$ in (5-6) suffices. However, the critical case $2N = n$ requires the stronger estimates $o(r^n)$.

independent proof of this identity. The following result contains the relevant details and some further information.

Theorem 5.6. *Let $k \in \mathbb{N}$. The operator $\iota^* S_k\big(g_+; \frac{n-k}{2}\big)$ defines a tangential differential operator $\mathcal{P}_k : C^\infty(M) \to C^\infty(M)$, i.e.,*

$$\iota^* S_k\Big(g_+; \frac{n-k}{2}\Big) = \mathcal{P}_k \iota^*.$$

For $k \in \mathbb{N}$ with $k \le n$ for even n, the operator $\mathcal{P}_k$ only depends on h and is conformally covariant, i.e.,

$$\mathcal{P}_k(\hat{h}) = e^{-((n+k)/2)\varphi} \circ \mathcal{P}_k(h) \circ e^{((n-k)/2)\varphi}$$

for $\hat{h} = e^{2\varphi} h$. For $k = 2N-1$, the operator $\mathcal{P}_k(h)$ vanishes identically. For $k = 2N$, the operator $\mathcal{P}_k(h)$ is proportional to the GJMS operator $P_{2N}(h)$ of (M,h):

$$\mathcal{P}_{2N}(h) = ((2N-1)!!)^2 P_{2N}(h). \tag{5-11}$$

Finally, we have

$$\iota^* S_{2N}\Big(g_+; \frac{n-1}{2} - N\Big) = (2N)!\,\iota^* P_{2N}(\bar{g}) \tag{5-12}$$

and

$$\iota^* S_{2N+1}\Big(g_+; \frac{n-3}{2} - N\Big) = (2N+2)!\,\iota^* \partial_r P_{2N}(\bar{g}). \tag{5-13}$$

Proof. We recall that a differential operator $D : C^\infty(M_+) \to C^\infty(M_+)$ restricts to a tangential operator with respect to M if and only if $D(rf) = rD'(f)$ for all $f \in C^\infty(M_+)$ and some differential operator $D' : C^\infty(M_+) \to C^\infty(M_+)$. By Lemma 4.6, we have

$$S_k(g_+; \lambda)(rf) = r S_k(g_+; \lambda - 1) f - k(2\lambda - n + k) S_{k-1}(g_+; \lambda - 1) f$$

for all $f \in C^\infty(M_+)$. Hence the Taylor series of $\iota^* S_k\big(g_+; \frac{n-k}{2}\big)(rf)$ in the variable r has vanishing constant term. It follows that $\iota^* S_k\big(g_+; \frac{n-k}{2}\big)$ defines a tangential operator, i.e., there is an operator

$$\mathcal{P}_k : C^\infty(M) \to C^\infty(M)$$

so that

$$\iota^* S_k\Big(g_+; \frac{n-k}{2}\Big) = \mathcal{P}_k \iota^*.$$

Now assume that $k \in \mathbb{N}$ with $k \le n$ for even n. The further properties of $\mathcal{P}_k$ follow from the relation between shift operators and residue families (Corollary 4.2). In particular, for these values of k, the operators $\mathcal{P}_k$ are determined by h and the conformal transformation law for $\mathcal{P}_k$ follows from Corollary 4.5. The fact that $\mathcal{P}_k$

vanishes identically for odd $k = 2N - 1$ is obvious. Indeed, by Corollary 4.2, we have

$$\begin{aligned}\mathcal{P}_{2N-1} &= \iota^* S_{2N-1}\Big(g_+;\, \frac{n+1}{2} - N\Big)\\ &= 2(-2N+1)_N(-N+1)_N D^{\mathrm{res}}_{2N-1}\Big(h;\, N - \frac{n+1}{2}\Big) = 0\end{aligned}$$

since $(-N+1)_N = 0$ and $D^{\mathrm{res}}_{2N-1}(h;\lambda)$ is regular in λ. By Corollary 4.2 and the first factorization relation in (5-1), we conclude that

$$\begin{aligned}\mathcal{P}_{2N}\iota^* &= \iota^* S_{2N}\Big(g_+;\, \frac{n}{2} - N\Big)\\ &= (-2N)_N\big(-N+\tfrac{1}{2}\big)_N D^{\mathrm{res}}_{2N}\Big(h;\, N - \frac{n}{2}\Big)\\ &= ((2N-1)!!)^2 P_{2N}(h)\iota^*.\end{aligned}$$

The identity (5-12) follows from Corollary 4.2 and the second factorization identity in (5-1). Similarly, the identity (5-13) follows from the factorization identity

$$D^{\mathrm{res}}_{2N+1}\Big(h;\, -\frac{n+1}{2} + N\Big) = D^{\mathrm{res}}_1\Big(h;\, -\frac{n+1}{2} - N\Big) P_{2N}(\bar{g})$$

for odd-order residue families (Theorem 5.4). By Corollary 4.2, this relation is equivalent to

$$\begin{aligned}\frac{1}{2(-2N-1)_{N+1}(-N-1)_{N+1}}\,&\iota^* S_{2N+1}\Big(g_+;\, -N + \frac{n-3}{2}\Big)\\ &= \frac{1}{2(N+1)}\iota^* S_1\Big(g_+;\, -N + \frac{n-3}{2}\Big) P_{2N}(\bar{g}).\end{aligned}$$

Simplification proves the claim. The proof is complete. □

Some comments on the latter results are in order.

Theorem 5.6 overlaps with [Gover and Waldron 2014, Theorems 4.1 and 4.5]. Indeed, assume that g_+ is Einstein. Then, by Proposition 3.2, $S_k(g_+;\lambda)$ can be regarded as a composition of k degenerate Laplacians $I \cdot D$. The fact that $\iota^* S_k\big(g_+;\, \frac{n-k}{2}\big)$ is tangential, is a consequence of a basic $\mathfrak{sl}(2)$-structure for the degenerate Laplacian [Gover and Waldron 2014, Section 3.1]. It holds true for compositions of degenerate Laplacians in a much more general setting. In the present situation, Lemmas 4.6 and 5.3 reflect that structure. In order to relate the tangential operator $\mathcal{P}_{2N}$ to the GJMS operator $P_{2N}(h)$, we used the first relation in (5-1). Note that this relation is a consequence of the basic residue relation (2-23) (derived in [Graham and Zworski 2003] from the ambient metric construction of $P_{2N}(h)$). In [Gover and Waldron 2014], the identification (5-11) also rests on ambient metric arguments.

As a consequence of Theorem 5.6, we obtain a new proof of the first system (5-2) of factorization identities for residue families.

Corollary 5.7. *Let $N \in \mathbb{N}$ with $N \leq n+1$ for even n. Then we have the factorization identities*

$$D_N^{\mathrm{res}}\Big(h; -\frac{n}{2}+N-k\Big) = P_{2k}(h) D_{N-2k}^{\mathrm{res}}\Big(h; -\frac{n}{2}+N-k\Big) \tag{5-14}$$

for $0 \leq 2k \leq N$.

Proof. The obvious relation

$$S_N(g_+; \lambda) = S_{2k}(g_+; \lambda) S_{N-2k}(g_+; \lambda + 2k)$$

implies the identity

$$\iota^* S_N\Big(g_+; \frac{n}{2}-k\Big) = \iota^* S_{2k}\Big(g_+; \frac{n}{2}-k\Big) S_{N-2k}\Big(g_+; \frac{n}{2}+2k\Big).$$

By Theorem 5.6, it is equivalent to

$$\iota^* S_N\Big(g_+; \frac{n}{2}-k\Big) = ((2k-1)!!)^2 P_{2k}(h) \iota^* S_{N-2k}\Big(g_+; \frac{n}{2}+2k\Big).$$

Now the identity (4-7) shows that, for even N, this relation can be restated as

$$(-2N)_N\big(-k+\tfrac{1}{2}\big)_N D_{2N}^{\mathrm{res}}\Big(h; -\frac{n}{2}+2N-k\Big)$$
$$= ((2k-1)!!)^2 (-2N+2k)_{N-k}\big(k+\tfrac{1}{2}\big)_{N-k} P_{2k}(h) D_{2N-2k}^{\mathrm{res}}\Big(h; -\frac{n}{2}+2N-k\Big).$$

Simplification proves the claim for even-order residues families. We omit the analogous proof for odd-order residue families which utilizes the identity (4-8). □

Since the family $S_N(g_+; \lambda)$ is a polynomial of degree N in λ, the N identities in Corollary 5.2 do *not* suffice to determine $S_N(g_+; \lambda)$ in terms of lower-order $S_k(g_+; \lambda)$ and GJMS operators of $\bar{g}$. However, that is possible by combining Corollary 5.2 with the following formula for the leading coefficient of that polynomial. We also recall that $w(r) = \sqrt{v(r)}$.

Proposition 5.8. *Let $N \in \mathbb{N}$. Then*

$$\frac{1}{N!} \frac{d^N}{d\lambda^N} S_N(g_+; \lambda) = (-2)^N w^{-1} \partial_r^N (w \cdot). \tag{5-15}$$

Proof. The relations (2-16) and

$$w^{-1} \partial_r (wf) = \tfrac{1}{2} v^{-1} \partial_r(v) f + \partial_r f, \quad f \in C^\infty(M_+^\circ)$$

show that we can rewrite the operator $S(g_+; \lambda)$ as

$$S(g_+; \lambda) = r \Delta_{\bar{g}} + (n-1)(\partial_r + v^{-1} \partial_r(v)) - 2\lambda w^{-1} \partial_r (w \cdot).$$

Now the leading coefficient of the polynomial $\lambda \to S_N(g_+; \lambda)$ coincides with the product of the leading coefficients of the N factors. But these are all given by $-2w^{-1}\partial_r(w \cdot)$. This proves the assertion. □

For later use, we introduce the notation $\partial_r^w \stackrel{\text{def}}{=} w^{-1}\partial_r(w\,\cdot)$. In these terms, the right-hand side of (5-15) equals $(-2)^N(\partial_r^w)^N$. In Section 6F, we shall discuss further consequences of Corollary 5.2 and Proposition 5.8.

Finally, we combine Proposition 5.8 with Corollary 4.2 to read off the leading coefficients of residue families. The definition of residue families implies formulas for these coefficients in terms of solution operators $\mathcal{T}_{2j}(h;\lambda)$ and renormalized volume coefficients v_{2j}. The following result shows how these can be simplified.

Corollary 5.9. *Let $N \in \mathbb{N}$ with $2N \le n$ for even n. The leading coefficient of the even-order residue family $\lambda \mapsto D_{2N}^{\text{res}}(h;\lambda)$ equals*

$$(-1)^N 2^{2N} \frac{N!}{(2N)!} \iota^* \partial_r^{2N}(w\,\cdot).$$

Similarly, the leading coefficient of the odd-order residue family $\lambda \mapsto D_{2N+1}^{\text{res}}(h;\lambda)$ equals

$$(-1)^N 2^{2N} \frac{N!}{(2N+1)!} \iota^* \partial_r^{2N+1}(w\,\cdot).$$

Proof. The even-order residue family $D_{2N}^{\text{res}}(h;\lambda)$ is a polynomial of degree N. By Proposition 5.8, Corollary 4.2 and $\iota^* w(r) = 1$, the coefficient of λ^N equals

$$(-2)^{2N} \frac{1}{(-2N)_N} \iota^* \partial_r^{2N}(w\,\cdot).$$

This proves the assertion. The proof in the odd-order case is analogous. ☐

5B. ***Shift operators and solution operators.*** In the present section, we assume that g_+ is Einstein and that n is odd. By the latter assumption, all Taylor coefficients of h_r are determined by h. The obvious modifications for even n are left to the reader. We relate the solution operator $\mathcal{T}_{2N}(h;\lambda)$ (see Section 2B) to the coefficients in the formal power series

$$S(g_+;\lambda) = -(2\lambda-n+1)\partial_r + r\sum_{k\ge 0} r^k S^{(k)}(h;\lambda) \tag{5-16}$$

following from the formal power series

$$\Delta_{\bar g} + (\lambda-n+1)\mathrm{J}(\bar g) = \sum_{k\ge 0} r^k S^{(k)}(h;\lambda).$$

Here we used the identity (3-5).[9] Only the operator $S^{(0)}(h;\lambda)$ contains two derivatives in r. By (2-19) and [Juhl 2009, Lemma 6.11], the first few coefficients in the

[9] The assumptions guarantee that (3-5) is an identity of formal power series.

expansion (5-16) are given by the operators

$$
\begin{aligned}
S^{(0)}(h;\lambda)f &= \Delta_h f + \partial_r^2 f + (\lambda-n+1)\mathsf{J}(h)f,\\
S^{(1)}(h;\lambda)f &= -\mathsf{J}(h)\partial_r f,\\
S^{(2)}(h;\lambda)f &= -\delta_h(\mathsf{P}(h)\#df) - \tfrac{1}{2}h(d\mathsf{J}(h),df) + \tfrac{1}{2}(\lambda-n+1)|\mathsf{P}(h)|^2 f,\\
S^{(3)}(h;\lambda)f &= -\tfrac{1}{2}|\mathsf{P}(h)|^2\partial_r f.
\end{aligned}
\tag{5-17}
$$

We recall that, under the present assumptions, the solution operators $\mathcal{T}_{2N}(h;\lambda)$ are well-defined for all $N\in\mathbb{N}$. In the following, we shall regard functions in $C^\infty(M)$ as functions on M_+ that do not depend on r, i.e., ∂_r annihilates functions in $C^\infty(M)$.

Proposition 5.10. *Let $N\in\mathbb{N}$. Then*

$$
-2N(2\lambda-n+2N)\mathcal{T}_{2N}(h;\lambda) = \sum_{k=0}^{N-1} S^{(2N-2k-2)}(h;n-\lambda-2k-1)\mathcal{T}_{2k}(h;\lambda)
\tag{5-18}
$$

as an identity of operators acting on $C^\infty(M)$. In particular, the $2N$-th GJMS operator on (M^n,h) is given by

$$
\begin{aligned}
&P_{2N}(h)\\
&\quad = -2^{2N-2}((N-1)!)^2\sum_{k=0}^{N-1} S^{(2N-2k-2)}\Big(h;\frac{n}{2}+N-2k-1\Big)\mathcal{T}_{2k}\Big(h;\frac{n}{2}-N\Big).
\end{aligned}
\tag{5-19}
$$

Proof. In (3-9), we replace the eigenfunction u by its asymptotic expansion

$$
\sum_{j\geq 0} r^{\nu+2j}a_{2j}(h;\nu) + \sum_{j\geq 0} r^{n-\nu+2j}b_{2j}(h;\nu)
$$

(see (2-20)). In order to simplify the following equations, we shall suppress the second sum. Then Theorem 3.7 implies

$$
\begin{aligned}
&S(g_+;\lambda)\Big(\sum_{j\geq 0} r^{\lambda+\nu-n+2j+1}a_{2j}(h;\nu)\Big)\\
&\qquad\qquad = (\lambda+\nu-n+1)(\nu-\lambda-1)\Big(\sum_{j\geq 0} r^{\lambda+\nu-n+2j}a_{2j}(h;\lambda)\Big).
\end{aligned}
$$

By Lemma 3.13, this relation is equivalent to

$$
\begin{aligned}
&\sum_{j\geq 0} r^{\lambda+\nu-n+2j+1}S(g_+;n-\nu-2j-1)a_{2j}(h;\nu)\\
&\qquad\qquad = -\sum_{j\geq 0} 2j(2\nu-n+2j)r^{\lambda+\nu-n+2j}a_{2j}(h;\nu).
\end{aligned}
$$

We rewrite this identity in terms of the expansion (5-16) and compare coefficients of powers of r. This gives

$$S^{(2j-2)}(h;n-\nu-1)a_0(h;\nu)+\cdots+S^{(0)}(h;n-\nu-2j+1)a_{2j-2}(h;\nu) = -2j(2\nu-n+2j)a_{2j}(h;\nu)$$

for $j \geq 1$. Using $a_{2k}(h;\nu) = \mathcal{T}_{2k}(h;\nu)a_0(h;\nu)$, we obtain

$$-2j(2\nu-n+2j)\mathcal{T}_{2j}(h;\nu) = \sum_{k=0}^{j-1} S^{(2j-2k-2)}(h;n-\nu-2k-1)\mathcal{T}_{2k}(h;\nu).$$

Note that $\mathcal{T}_{2j}(h;\nu)$ has a simple pole at $\nu = \frac{n}{2} - j$ with residue given by $P_{2j}(h)$ (see (2-23)). Thus the last display implies

$$P_{2j}(h) = -2^{2j-2}((j-1)!)^2 \sum_{k=0}^{j-1} S^{(2j-2k-2)}\Big(h;\frac{n}{2}+j-2k-1\Big)\mathcal{T}_{2k}\Big(h;\frac{n}{2}-j\Big).$$

The proof is complete. □

Proposition 5.10 is a compressed version of the usual algorithm for the calculation of the solution operators. We shall illustrate it by low-order examples in Section 6.

5C. ***Holographic formulas for Q-curvatures.*** In the present section, we assume that n is even. We shall discuss new holographic formulas for the Q-curvatures $Q_{2N}(h)$ for $2N \leq n$.

We start with a simple proof of a result which is also of independent interest (see [Baum and Juhl 2010, Theorem 1.6.6]). It has been useful in connection with a discussion of the recursive structure of Q-curvatures [Juhl 2014].

Proposition 5.11. *Assume that n is even and let $N \in \mathbb{N}$ with $2N \leq n$. Then*

$$D_{2N}^{\text{res}}(h;0)(1) = 0.$$

Proof. Corollary 4.2 implies that

$$D_{2N}^{\text{res}}(h;0)(1) = \frac{1}{(-2N)_N\left(\frac{n+1}{2}-2N\right)_N}\Sigma_{2N}(h;n-2N)(1).$$

By definition, we have

$$S_{2N}(g_+;n-2N) = S(g_+;n-2N)\circ\cdots\circ S(g_+;n-1).$$

But (3-1) shows that

$$S(g_+;n-1)(1) = 0. \tag{5-20}$$

Hence $\Sigma_{2N}(h;n-2N)(1) = 0$. This completes the proof. □

The polynomial $\lambda \mapsto D_{2N}^{\mathrm{res}}(h;\lambda)(1)$ is called the Q-curvature polynomial.

The Q-curvature polynomial $D_{2N}^{\mathrm{res}}(h;\lambda)(1)$ also vanishes in odd dimensions n. The above arguments prove this fact if $\left(\frac{n+1}{2}-2N\right)_N \neq 0$.[10]

In the critical case $2N = n$, Proposition 5.11 states that $D_n^{\mathrm{res}}(h;0)(1) = 0$. Of course, this result also follows from $D_n^{\mathrm{res}}(h;0) = P_n(h)\iota^*$ (see (5-1)) and $P_n(h)(1) = 0$.

We continue with the discussion of the critical Q-curvature $Q_n(h)$ and recall the holographic formula [Juhl 2009, Theorem 6.6.1]

$$Q_n(h) = -(-1)^{n/2}\dot{D}_n^{\mathrm{res}}(h;0)(1). \tag{5-21}$$

The following result is a consequence of this identity.

Theorem 5.12 (holographic formula for critical Q-curvature). *Let n be even. Then*

$$Q_n(h) = (-1)^n/2D_{n-1}^{\mathrm{res}}(h;-1)\partial_r(\log v) \tag{5-22}$$

or, equivalently,

$$Q_n(h) = c_n\Sigma_{n-1}(h;0)\partial_r(\log v) \tag{5-23}$$

with $c_n = (-1)^{n/2}2^{n-2}\left(\Gamma\left(\frac{n}{2}\right)/\Gamma(n)\right)^2$.

We recall that the composition $\Sigma_{n-1}(h;0) = \iota^* S_{n-1}(g_+;0)$ only depends on h.

Proof. The critical special case $2N = n$ of the first factorization identity in Corollary 4.3 reads

$$D_n^{\mathrm{res}}(h;\lambda) = D_{n-1}^{\mathrm{res}}(h;\lambda-1)S(g_+;\lambda+n-1).$$

Now we differentiate this relation at $\lambda = 0$ and use the vanishing result (5-20). We obtain

$$\dot{D}_n^{\mathrm{res}}(h;0)(1) = D_{n-1}^{\mathrm{res}}(h;-1)\dot{S}(g_+;n-1)(1)$$

But Proposition 5.8 shows that $\dot{S}(g_+;n-1) = -2w^{-1}\partial_r(w\,\cdot)$. Hence

$$\dot{S}(g_+;n-1)(1) = -2w^{-1}\partial_r(w) = -2\partial_r(\log w) = -\partial_r(\log v).$$

The relation (5-22) follows by combining these facts with the holographic formula (5-21). The second relation (5-23) follows by combining the first relation with Corollary 4.2. □

Remark 5.13. Formula (5-23) should be compared with the special case

$$((n-1)!!)^2 Q_n(h) = \iota^*(I\cdot D)_{n-1}[-n+1] \circ (I\cdot D)_L[1]\log(1) \tag{5-24}$$

[10] In the remaining cases, the polynomial $\Sigma_{2N}(h;\lambda)(1)$ has a double zero at $\lambda = n-2N$.

of [Gover and Waldron 2014, Theorem 4.7], where we set $(I{\cdot}D)[\lambda] = (I{\cdot}D)[\bar{g}, r; \lambda]$ and likewise for $(I{\cdot}D)_L$. Here the factor $(I{\cdot}D)_L$ is defined to act on the log density $\log(\mu)$ (μ any positive smooth function) according to

$$(I{\cdot}D)_L[g,\sigma;\omega]\log(\mu) \\ \stackrel{\text{def}}{=} \big[-\sigma\Delta_g + (n-1)g(d\sigma, d\cdot)\big]\log(\mu) - \frac{\omega}{n+1}\big[(n-1)\Delta_g(\sigma) + 2n\sigma\mathsf{J}(g)\big]$$

(see [Gover and Waldron 2014, Section 2]). For $g = \bar{g}$, $\sigma = r$ and $\mu = 1$, we find

$$\begin{aligned}(I{\cdot}D)_L[\bar{g}, r;\omega]\log(1) &= -\frac{\omega}{n+1}\big[(n-1)\Delta_{\bar{g}}(r) + 2nr\mathsf{J}(\bar{g})\big] \\ &= -\frac{\omega}{n+1}\Big[\frac{n-1}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r) - n\operatorname{tr}(h_r^{-1}\dot{h}_r)\Big] \\ &= \frac{\omega}{2}\operatorname{tr}(h_r^{-1}\dot{h}_r) = \omega\partial_r(\log v).\end{aligned} \tag{5-25}$$

Hence the formula (5-24) reduces to (5-23), up to a sign due to conventions. Now the transformation law $\mu^n Q_n(\mu^2 h) = P_n(h)\log(\mu) + Q_n(h)$ and (5-11) imply

$$((n-1)!!)^2\mu^n Q_n(\mu^2 h) \\ = \iota^*(I{\cdot}D)_{n-1}[-n+1]\circ(I{\cdot}D)[0]\log(\mu) + \iota^*(I{\cdot}D)_{n-1}[-n+1]\circ(I{\cdot}D)_L[1]\log(1).$$

But the relation

$$(I{\cdot}D)_L[1]a - (I{\cdot}D)_L[1]b = (I{\cdot}D)[0](a-b)$$

yields

$$(I{\cdot}D)[0]\log(\mu) + (I{\cdot}D)_L[1]\log(1) = (I{\cdot}D)_L[1]\log(\mu).$$

Hence

$$((n-1)!!)^2\mu^n Q_n(\mu^2 h) = \iota^*(I{\cdot}D)_{n-1}[-n+1](I{\cdot}D)_L[1]\log(\mu).$$

This proves [Gover and Waldron 2014, Theorem 4.7].[11]

The formula (5-23) resembles the well-known beautiful formula

$$Q_n(h) = (-1)^{n/2-1}\boldsymbol{\Delta}^{n/2}(\log t)|_{\rho=0,\, t=1}$$

of Fefferman and Hirachi [2003]. Here $\boldsymbol{\Delta}$ denotes the Laplacian of the ambient metric in normal form relative to h, and t is the homogeneous coordinate on the ambient space $\mathbb{R}^+ \times M \times (-\varepsilon, \varepsilon)$ with coordinates (t, x, ρ).

Next, we establish a generalization of Theorem 5.12 to all subcritical Q-curvatures.

[11] Strictly speaking, μ on the right-hand side is a function on M_+ and the formula for Q_n does only depend on the restriction of μ to the hypersurface M.

Theorem 5.14 (holographic formula for subcritical Q-curvatures). *Let n be even and assume that $2N < n$. Then*

$$Q_{2N}(h) = c_{2N}\,\Sigma_{2N-1}\Big(h;\,\frac{n}{2}-N\Big)\partial_r(\log v), \tag{5-26}$$

where $c_{2N} = (-1)^N 2^{2N-2}(\Gamma(N)/\Gamma(2N))^2$. *Equivalently,*

$$Q_{2N}(h) = (-1)^N D^{\mathrm{res}}_{2N-1}\Big(h;\,-\frac{n}{2}+N-1\Big)\partial_r(\log v). \tag{5-27}$$

Proof. On the one hand, we have

$$D^{\mathrm{res}}_{2N}\Big(h;\,-\frac{n}{2}+N\Big)(1) = P_{2N}(h)(1) = (-1)^N\Big(\frac{n}{2}-N\Big)Q_{2N}(h)$$

using (2-5) and (5-1). On the other hand, Corollary 4.2 implies

$$D^{\mathrm{res}}_{2N}\Big(h;\,-\frac{n}{2}+N\Big)(1) = \frac{1}{(-2N)_N\big(\frac{1}{2}-N\big)_N}\Sigma_{2N}\Big(h;\,\frac{n}{2}-N\Big)(1).$$

But

$$\Sigma_{2N}\Big(h;\,\frac{n}{2}-N\Big) = \Sigma_{2N-1}\Big(h;\,\frac{n}{2}-N\Big)S\Big(g_+;\,\frac{n}{2}+N-1\Big)$$

and

$$S\Big(g_+;\,\frac{n}{2}+N-1\Big)(1) = \Big(\frac{n}{2}-N\Big)\partial_r(\log v).$$

By comparing both expressions for $D^{\mathrm{res}}_{2N}\big(h;\,-\frac{n}{2}+N\big)(1)$, we obtain

$$Q_{2N}(h) = (-1)^N\frac{1}{(-2N)_N\big(\frac{1}{2}-N\big)_N}\Sigma_{2N-1}\Big(h;\,\frac{n}{2}-N\Big)\partial_r(\log v).$$

Now simplification proves the first assertion. The second follows from this result using the second relation in Corollary 4.2. □

Theorem 5.12 obviously follows from Theorem 5.14 by analytic continuation in the dimension n. However, the above proof avoids this argument.

Finally, we prove that the formulas (5-23) and (5-27) are equivalent to the well-known holographic formulas for Q-curvatures proved in [Graham and Juhl 2007; Juhl 2011]. In fact, the definitions imply that (5-27) is equivalent to

$$(-1)^N Q_{2N} = 2^{2N-2}(N-1)!^2 \times\bigg(\sum_{j=0}^{2N-1}\mathcal{T}_j^*\big(\tfrac{n}{2}-N\big)v_0+\cdots+\mathcal{T}_0^*\big(\tfrac{n}{2}-N\big)v_j\bigg)\frac{1}{(2N-1-j)!}\iota^*\partial_r^{2N-1-j}\Big(\frac{\dot v}{v}\Big).$$

In the latter sum, the operator $\mathcal{T}_{2k}^*\big(\frac{n}{2}-N\big)$ acts on

$$\bigg(v_0\frac{1}{(2N-1-2k)!}\iota^*\partial_r^{2N-1-2k}+\cdots+v_{2N-2k-2}\iota^*\partial_r\bigg)\Big(\frac{\dot v}{v}\Big).$$

But this sum equals the $(2N-1-2k)$-th Taylor coefficient of $v(\dot{v}/v) = \dot{v}$, i.e., equals $(2N-2k)v_{2N-2k}$. Hence the above formula simplifies to

$$(-1)^N Q_{2N} = 2^{2N-2}(N-1)!^2 \sum_{k=0}^{N-1} (2N-2k)\mathcal{T}_{2k}^*\big(\tfrac{n}{2}-N\big)(v_{2N-2k}).$$

This formula is equivalent to [Juhl 2011, Theorem 1.1]. The same arguments also apply in the critical case. This completes the proof.

In summary, the above discussion provides reformulations and easy new proofs of well-known holographic formulas for Q-curvatures (in even dimension). The key arguments here are the second identity in Corollary 4.2 and/or the first identity in Corollary 4.3.

6. A panorama of examples

In the present section, we illustrate the main results (Theorems 4.1, 5.1, 5.6, 5.12 and Proposition 5.10) by low-order examples. Furthermore, we discuss certain remarkable expansion of the families $S_N(g_+;\lambda)$ with respect to the parameter r.

We shall often simplify notation by omitting the obvious metrics. In particular, we shall write Δ for Δ_h, J for $\mathsf{J}(h)$, P for $\mathsf{P}(h)$, $S(\lambda)$ for $S(g_+;\lambda)$ etc.

We also recall the expansion [Juhl 2009, Section 6.11]

$$\Delta_{\bar{g}} f = [\Delta f + \partial_r^2 f] - r\mathsf{J}\partial_r f + r^2\big[-\delta(\mathsf{P}\#df) - \tfrac{1}{2}(d\mathsf{J}, df)\big] + O(r^3). \tag{6-1}$$

By $\dot{v}/v = 2rv_2 + O(r^3)$ with $v_2 = -\frac{1}{2}\mathsf{J}(h)$, we have

$$S(g_+;\lambda) = -(2\lambda-n+1)\partial_r + r\big[\Delta_h + \partial_r^2 + (\lambda-n+1)\mathsf{J}(h))\big] + O(r^2).$$

6A. ***Theorem 4.1 for N ≤ 3.*** The first-order family $\delta_1(\lambda)$ is given by [Juhl 2009, Section 6.2]

$$\delta_1(\lambda) = \iota^*\partial_r.$$

On the other hand, we have

$$\Sigma_1(\lambda) = \iota^* S_1(\lambda) = -(2\lambda-n+1)\iota^*\partial_r = -(2\lambda-n+1)\delta_1(\lambda). \tag{6-2}$$

This confirms Theorem 4.1 in the case $N=1$.

The second-order family $\delta_2(\lambda)$ is given by [Juhl 2009, Section 6.7]

$$\delta_2(\lambda) = \tfrac{1}{2}\iota^*\partial_r^2 + \frac{1}{2(n-2-2\lambda)}(\Delta + (\lambda-n+2)\mathsf{J})\iota^*.$$

On the other hand, by definition, we have

$$\begin{aligned}\Sigma_2(\lambda) &= \iota^* S_2(\lambda) = \iota^* S(\lambda)S(\lambda+1)\\ &= -(2\lambda-n+1)\iota^*\partial_r\big[r\Delta_{\bar{g}} + (-2\lambda+n-3)\partial_r + (\lambda-n+2)r\mathsf{J}\big].\end{aligned}$$

By (6-1), this formula simplifies to

$$
\begin{aligned}
\text{(6-3)}\quad \Sigma_2(\lambda) &= -(2\lambda-n+1)\big[\Delta\iota^* - (2\lambda-n+2)\iota^*\partial_r^2 + (\lambda-n+2)\mathsf{J}\iota^*\big] \\
&= (-2)_2(2\lambda-n+1)_2\Big[\tfrac{1}{2}\iota^*\partial_r^2 - \frac{1}{2(2\lambda-n+2)}(\Delta\iota^* + (\lambda-n+2)\mathsf{J}\iota^*)\Big] \\
&= (-2)_2(2\lambda-n+1)_2\delta_2(\lambda).
\end{aligned}
$$

This identity confirms Theorem 4.1 for $N = 2$.

Note that, for $n = 3$, the latter formula gives

$$\Sigma_2(\lambda)(1) = -(-2)_2(2\lambda-2)_2\frac{(\lambda-1)}{2(2\lambda-1)}\mathsf{J} = -2(\lambda-1)^2\mathsf{J}.$$

This confirms the double zero mentioned after Proposition 5.11. In this case, both factors in the definition of $\Sigma_2(\lambda)$ contribute a zero.

The third-order family $\delta_3(\lambda)$ is given by [Juhl 2009, Section 6.8]

$$\delta_3(\lambda) = \tfrac{1}{6}\iota^*\partial_r^3 + \frac{1}{2(n-2-2\lambda)}(\Delta + (\lambda-n+2)\mathsf{J})\iota^*\partial_r.$$

On the other hand, by definition, we have

$$
\begin{aligned}
\Sigma_3(\lambda) = \iota^* S_3(\lambda) &= \iota^* S(\lambda)S(\lambda+1)S(\lambda+2) \\
&= -(2\lambda-n+1)\iota^*\partial_r\big[r\Delta_{\bar g} - (2\lambda-n+3)\partial_r + (\lambda-n+2)r\mathsf{J}\big] \\
&\qquad \circ \big[r\Delta_{\bar g} - (\lambda-n+5)\partial_r + (\lambda-n+3)r\mathsf{J}\big].
\end{aligned}
$$

By (6-1), the last display simplifies to

$$
\begin{aligned}
\text{(6-4)}\quad & -(2\lambda-n+1)\Big[(-(\lambda-n+5) + 2 - 2(2\lambda-n+3))\Delta\iota^*\partial_r \\
&\quad + (2-(2\lambda-n+5)-2(2\lambda-n+3)+(2\lambda-n+3)(2\lambda-n+5))\iota^*\partial_r^3 \\
&\quad + (-2+2(\lambda-n+3) + 2(2\lambda-n+3) - 2(2\lambda-n+3)(\lambda-n+3) \\
&\quad - (\lambda-n+2)(2\lambda-n+5))\mathsf{J}\iota^*\partial_r\Big] \\
&= (-3)_3(2\lambda-n+1)_3\Big[\tfrac{1}{6}\iota^*\partial_r^3 - \frac{1}{2(2\lambda-n+2)}\big[\Delta\iota^*\partial_r + (\lambda-n+2)\mathsf{J}\iota^*\partial_r\big]\Big] \\
&= (-3)_3(2\lambda-n+1)_3\delta_3(\lambda).
\end{aligned}
$$

This identity confirms Theorem 4.1 for $N = 3$.

6B. ***Theorem 5.1 for N = 1.*** We have shown that the family $S_N(g_+;\lambda)$ contains the GJMS operator $P_{2N}(\bar g)$ if g_+ is an Einstein metric on M_+°. Now we give a direct proof for the first example. We recall that $m = \frac{n+1}{2}$. By formula (3-5), we compute

$$S_1(g_+; m-1) = S(g_+; m-1) = r\Big(\Delta_{\bar g} - \frac{n-1}{2}\mathsf{J}(\bar g)\Big)$$

if g_+ is Einstein. By (2-2), the right-hand side coincides with $r P_2(\bar{g})$. For general Poincaré metrics, the identity holds with error terms (see Remark 3.5).

6C. ***Theorem 5.6 for $N = 1$.*** It shows that the family $S_{2N}(g_+; \lambda)$ induces a tangential operator which is proportional to the GJMS operators for (M, h). Now we compute

$$\begin{aligned}\iota^* S_2\Big(\frac{n}{2}-1\Big) &= \iota^* S\Big(\frac{n}{2}-1\Big) S\Big(\frac{n}{2}\Big)\\ &= \iota^* \partial_r \Big[-\partial_r + r\Big[\Delta + \partial_r^2 - \Big(\frac{n}{2}-1\Big)\mathsf{J}\Big]\Big]\\ &= \Delta \iota^* - \Big(\frac{n-2}{2}\Big)\mathsf{J}\iota^*,\end{aligned}$$

which coincides with $P_2 \iota^*$ (see (2-2)).

Note also that

$$\iota^* S_2\Big(g_+; \frac{n-1}{2}-1\Big) = 2\iota^* \partial_r \Big[r \Delta_{\bar{g}} - \Big(\frac{n-1}{2}\Big) r \mathsf{J}\Big] = 2\iota^* P_2(\bar{g})$$

and

$$\iota^* S_3\Big(g_+; \frac{n-3}{2}-1\Big) = 4!\iota^* \partial_r P_2(\bar{g})$$

by (6-3) and (6-4) (for Einstein g_+ see also (6-8) and (6-11)). The latter two identities are special cases of (5-12) and (5-13).

6D. ***Proposition 5.10 for $N \le 2$.*** We recall that [Juhl 2009, Section 6.7]

$$\mathcal{T}_2(\lambda) = \frac{1}{2(n-2\lambda-2)}[\Delta - \lambda \mathsf{J}]. \tag{6-5}$$

By (5-16), we have

$$S^{(0)}(n-\lambda-1) = \Delta + \partial_r^2 - \lambda \mathsf{J}.$$

Hence we obtain

$$-2(2\lambda-n+2)\mathcal{T}_2(\lambda) = S^{(0)}(n-\lambda-1)\mathcal{T}_0(\lambda)$$

as an identity of operators on $C^\infty(M)$. This coincides with (5-18) for $N = 1$.

We recall that [Juhl 2009, Theorem 6.9.4]

$$\begin{aligned}&\mathcal{T}_4(\lambda)\\ &= \frac{1}{4(n-4-2\lambda)}\Big[\frac{1}{2(n-2-2\lambda)}[\Delta-(\lambda+2)\mathsf{J}][\Delta-\lambda\mathsf{J}] - \tfrac{1}{2}\lambda|\mathsf{P}|^2 - \delta(\mathsf{P}\#d) - \tfrac{1}{2}(d\mathsf{J}, d)\Big].\end{aligned}$$

Now, using (5-17) and (6-5), we compute

$$\begin{aligned}&S^{(2)}(n-\lambda-1)\mathcal{T}_0(\lambda) + S^{(0)}(n-\lambda-3)\mathcal{T}_2(\lambda)\\ &\qquad = -\delta(\mathsf{P}\#d) - \tfrac{1}{2}(d\mathsf{J}, d) - \tfrac{1}{2}|\mathsf{P}|^2 + \frac{1}{2(n-2\lambda-2)}[\Delta - (\lambda+2)\mathsf{J}][\Delta - \lambda\mathsf{J}].\end{aligned}$$

Hence

$$-4(2\lambda-n+4)\mathcal{T}_4(\lambda) = S^{(2)}(n-\lambda-1)\mathcal{T}_0(\lambda) + S^{(0)}(n-\lambda-3)\mathcal{T}_2(\lambda).$$

This coincides with (5-18) for $N = 2$.

6E. ***Theorem 5.12 for n ≤ 4 and Theorem 5.14 for N ≤ 2.*** We first confirm (5-23) for $n = 2$ and $n = 4$. As a preparation, we note that

$$\partial_r(\log v) = \dot{v}(r)/v(r) = 2rv_2 + r^3(4v_4 - 2v_2^2) + \cdots .$$

For $n = 2$, (5-23) claims that

$$Q_2 = -\Sigma_1(0)\partial_r(\log v).$$

Now, using (6-2), this identity simplifies to

$$Q_2 = -2v_2.$$

For $n = 4$, (5-23) claims that

$$Q_4 = \tfrac{4}{36}\Sigma_3(0)\partial_r(\log v).$$

Using (6-4), this formula reads

$$Q_4 = 4\big[\tfrac{1}{6}\iota^*\partial_r^3 + \tfrac{1}{4}(\Delta\iota^*\partial_r - 2\mathsf{J}\iota^*\partial_r)\big]\partial_r(\log v).$$

The latter expression simplifies to

$$16v_4 - 8v_2^2 + 2(\Delta - 2\mathsf{J})v_2.$$

Now the standard formulas

$$v_2 = -\tfrac{1}{2}\mathsf{J} \quad \text{and} \quad v_4 = \tfrac{1}{8}(\mathsf{J}^2 - |\mathsf{P}|^2)$$

confirm that the above formulas are equivalent to the well-known expressions

$$Q_2 = \mathsf{J} \quad \text{and} \quad Q_4 = 2\mathsf{J}^2 - 2|\mathsf{P}|^2 - \Delta\mathsf{J}.$$

Similar calculations confirm (5-26) for $N = 1$ and $N = 2$. In fact, by (6-2), the assertion

$$Q_2 = -\Sigma_1\Big(\frac{n}{2} - 1\Big)\partial_r(\log v)$$

is equivalent $Q_2 = -2v_2$. Similarly, by (6-4), the assertion

$$Q_4 = \tfrac{4}{36}\Sigma_3\Big(\frac{n}{2} - 2\Big)\partial_r(\log v)$$

is equivalent to

$$Q_4 = 16v_4 - 8v_2^2 + 2\Delta v_2 - n\mathsf{J}v_2 = \frac{n}{2}\mathsf{J}^2 - 2|\mathsf{P}|^2 - \Delta\mathsf{J}.$$

These results reproduce the formulas in (2-6).

6F. ***Some interesting expansions of $S_N(g_+;\lambda)$.*** We finish this section with a discussion of some interesting expansions of the families $S_N(g_+;\lambda)$. Here we restrict to the low-order examples $N \le 3$. In order to simplify the presentation, we shall also assume that g_+ is Einstein. The expansions in question describe the families $S_N(g_+;\lambda)$ as polynomials in r with coefficients that are polynomials in the GJMS operators $P_{2k}(\bar{g})$ for $k \le N$ and the operators $\partial_r^w = w^{-1}\partial_r(w\cdot)$. The existence of such expansions follows from Theorem 5.1, Corollary 5.2 and Proposition 5.8. Of particular interest will be the coefficient of r^N in the expansion of $S_N(h;\lambda)$. This coefficient is a polynomial in λ the leading coefficient of which is related to the second-order operators $\mathcal{M}_2(\bar{g})$, $\mathcal{M}_4(\bar{g})$ and $\mathcal{M}_6(\bar{g})$ (see (6-9), (6-12)) associated to $(M_+,\bar{g})$. We recall that these operators are the first coefficients in the r-expansion of the holographic Laplacian introduced in [Juhl 2013]. These experiments provide evidence for a general property of these expansions which will be formulated at the end.

As in Theorem 5.1, we shall use the notation $m = \frac{n+1}{2}$.

Example 6.1. The second-order family $S_1(g_+;\lambda)$ is linear in the variable λ and satisfies two identities:

$$S_1(g_+;m-1) = r P_2(\bar{g}),$$
$$\frac{d}{d\lambda} S_1(g_+;\lambda) = -2\partial_r^w.$$

These identities imply the representation

$$S_1(g_+;\lambda) = -2(\lambda-m+1)\partial_r^w + r P_2(\bar{g}). \tag{6-6}$$

In particular, the coefficient of r is given by $\mathcal{M}_2(\bar{g}) = P_2(\bar{g})$.

Example 6.2. The fourth-order family $S_2(g_+;\lambda)$ is quadratic in the variable λ and satisfies three identities:

$$S_2(g_+;m-1) = r^2 P_4(\bar{g}),$$
$$S_2(g_+;m-2) = S_1(g_+;m-2) r P_2(\bar{g}),$$
$$\frac{1}{2!}\frac{d^2}{d\lambda^2} S_2(g_+;\lambda) = 4(\partial_r^w)^2.$$

Hence $S_2(g_+;\lambda)$ can be written in the form

$$S_2(g_+;\lambda) = 4(\lambda-m+1)_2(\partial_r^w)^2 - (\lambda-m+1) S_1(g_+;m-2) r P_2(\bar{g}) + (\lambda-m+2) r^2 P_4(\bar{g}). \tag{6-7}$$

Now applying Lemma 4.6 to the middle summand gives

$$S_1(g_+;m-2) r P_2(\bar{g}) = r S_1(g_+;m-3) P_2(\bar{g}) + 2P_2(\bar{g}).$$

Thus, by combination with (6-6), we can rewrite (6-7) as

$$S_2(g_+;\lambda) = 4(\lambda-m+1)_2(\partial_r^w)^2 - 2(\lambda-m+1) P_2(\bar{g}) - 4(\lambda-m+1) r \partial_r^w P_2(\bar{g}) + r^2\big[(\lambda-m+2) P_4(\bar{g}) - (\lambda-m+1) P_2(\bar{g})^2\big]. \tag{6-8}$$

In particular, the coefficient of r^2 is a linear polynomial in λ the leading coefficient of which is given by the operator

$$\mathcal{M}_4(\bar{g}) = P_4(\bar{g}) - P_2(\bar{g})^2. \tag{6-9}$$

Example 6.3. The sixth-order family $S_3(g_+;\lambda)$ is cubic in the variable λ and satisfies four identities:

$$\begin{aligned} S_3(g_+; m-1) &= r^3 P_6(\bar{g}), \\ S_3(g_+; m-2) &= S_1(g_+; m-2) r^2 P_4(\bar{g}), \\ S_3(g_+; m-3) &= S_2(g_+; m-3) r P_2(\bar{g}), \\ \frac{1}{3!}\frac{d^3}{d\lambda^3} S_3(g_+;\lambda) &= -8(\partial_r^w)^3. \end{aligned}$$

Hence $S_3(g_+;\lambda)$ can be represented in the form

$$\begin{aligned} S_3(g_+;\lambda) = -8(\lambda-m+1)_3(\partial_r^w)^3 + \tfrac{1}{2}(\lambda-m+1)_2 S_2(g_+; m-3) r P_2(\bar{g}) \\ - (\lambda-m+1)(\lambda-m+3) S_1(g_+; m-2) r^2 P_4(\bar{g}) \\ + \tfrac{1}{2}(\lambda-m+2)_2 r^3 P_6(\bar{g}). \end{aligned} \tag{6-10}$$

Applying Lemma 4.6 to the middle two summands yields the relations

$$\begin{aligned} S_2(g_+; m-3)(r P_2(\bar{g})) &= r S_2(g_+; m-4) P_2(\bar{g}) + 6 S_1(g_+; m-3) P_2(\bar{g}), \\ S_1(g_+; m-2)(r^2 P_4(\bar{g})) &= r^2 S_1(g_+; m-4) P_4(\bar{g}) + 6 r P_4(\bar{g}). \end{aligned}$$

Hence, by combination with the respective representations (6-6) and (6-8) of $S_1(g_+;\lambda)$ and $S_2(g_+;\lambda)$, we can rewrite $S_3(g_+;\lambda)$ in the form

$$\begin{aligned} &(2\lambda-n+1)(2\lambda-n+3)\big(3\partial_r^w P_2(\bar{g}) - (2\lambda-n+5)(\partial_r^w)^3\big) \\ &\quad - \tfrac{3}{2}(2\lambda-n+1) r \Big[(2\lambda-n+5) P_4(\bar{g}) - (2\lambda-n+3) P_2(\bar{g})^2 \\ &\qquad\qquad - 2(2\lambda-n+3)(\partial_r^w)^2 P_2(\bar{g})\Big] \\ &\quad - \tfrac{3}{2}(2\lambda-n+1) r^2 \big[(2\lambda-n+5)\partial_r^w P_4(\bar{g}) - (2\lambda-n+3)\partial_r^w P_2(\bar{g})^2\big] \\ &\quad + \tfrac{1}{8} r^3 \Big[-2(2\lambda-n+1)(2\lambda-n+5) P_2(\bar{g}) P_4(\bar{g}) + (2\lambda-n+3)(2\lambda-n+5) P_6(\bar{g}) \\ &\qquad -2(2\lambda-n+1)(2\lambda-n+3) P_4(\bar{g}) P_2(\bar{g}) + 3(2\lambda-n+1)(2\lambda-n+3) P_2(\bar{g})^3\Big]. \end{aligned} \tag{6-11}$$

In particular, the coefficient of r^3 is a quadratic polynomial in λ the leading coefficient of which is a constant multiple of

$$\mathcal{M}_6(\bar{g}) = P_6(\bar{g}) - 2P_2(\bar{g})P_4(\bar{g}) - 2P_4(\bar{g})P_2(\bar{g}) + 3P_2(\bar{g})^3. \tag{6-12}$$

The above examples suggest that in the analogous representation of $S_N(g_+;\lambda)$ the coefficient of r^N is a constant multiple of the degree $N-1$ polynomial

$$\sum_{|I|=N} m_I \frac{\left(\lambda-\frac{n-1}{2}\right)_N}{\lambda-\frac{n-1}{2}+N-I_r} P_{2I}(\bar{g}). \tag{6-13}$$

Here the sum runs over all partitions $I=(I_1,\dots,I_r)$ of size $|I|=N$ and, for any I, we set $P_{2I}=P_{2I_1}\cdots P_{2I_r}$; for more details on the coefficients m_I we refer to [Juhl 2013]. The expression (6-13) resembles the polynomials in [Juhl 2013, Theorem 4.1] which describes residue families in terms of GJMS operators. Note that the leading coefficient of the polynomial (6-13) is the building block operator

$$\mathcal{M}_{2N}(\bar{g})=\sum_{|I|=N} m_I P_{2I}(\bar{g})$$

(see Section 7). We will return to that problem in later work.

7. Epilogue

In the present section, we sketch some further developments and indicate some interesting future developments.

We first prove a consequence of the expansions discussed in Section 6F. Then we generalize this result to arbitrary order. We show that the result naturally follows from a basic property of an operator which in [Juhl 2016] was termed the holographic Laplacian. We expect that these generalizations play a similar role in the study of the families $S_N(g_+;\lambda)$ of arbitrary order $N\in\mathbb{N}$.

In order to simplify the presentations as much as possible, we assume that M is an analytic manifold of odd dimension n. For an analytic metric h on M, we let g_+ be a Poincaré metric in normal form relative to h. It satisfies $\operatorname{Ric}(g_+)+ng_+=0$ and, for any $N\in\mathbb{N}$, there is a well-defined GJMS operator $P_{2N}(h)$.

We also recall the notation $\partial_r^w=w^{-1}\partial_r(w\cdot)$ and

$$\mathcal{M}_2(h)=P_2(h),\quad \mathcal{M}_4(h)=P_4(h)-P_2(h)^2.$$

As usual, let $\bar{g}=r^2g_+$. The analogous definitions yield $\mathcal{M}_2(\bar{g})$ and $\mathcal{M}_4(\bar{g})$.

Theorem 7.1. $$r\mathcal{M}_4(\bar{g})=2[\partial_r^w,\mathcal{M}_2(\bar{g})].$$

Proof. Using (6-6), we compute

$$\begin{aligned}S_2(g_+;\lambda)&=S_1(g_+;\lambda)S_1(g_+;\lambda+1)\\&=S_1(g_+;\lambda)\big[-2(\lambda-m+2)\partial_r^w+rP_2(\bar{g})\big]\\&=-2(\lambda-m+2)S_1(g_+;\lambda)\partial_r^w+S_1(g_+;\lambda)rP_2(\bar{g}).\end{aligned}$$

Now we apply Lemma 3.13 to move the variable r in the last term to the left, i.e.,

$$S_2(g_+;\lambda) = -2(\lambda-m+2)S_1(g_+;\lambda)\partial_r^w + rS_1(g_+;\lambda)P_2(\bar g) - 2(\lambda-m+1)P_2(\bar g)).$$

By another application of (6-6), we conclude

$$\begin{aligned} S_2(g_+;\lambda) &= 2(\lambda-m+2)\big[-2(\lambda-m+1)(\partial_r^w)^2 + rP_2(\bar g)\partial_r^w\big] \\ &\qquad + r[-2(\lambda-m)\partial_r^w P_2(\bar g) + P_2(\bar g)^2 - 2(\lambda-m+1)P_2(\bar g) \\ &= 4(\lambda-m+1)_2(\partial_r^w)^2 - 2(\lambda-m+1)P_2(\bar g) \\ &\qquad - 2r\big[(\lambda-m+2)P_2(\bar g)\partial_r^w + (\lambda-m)\partial_r^w P_2(\bar g)\big] + r^2P_2(\bar g)^2. \end{aligned}$$

The difference between (6-8) and the last display gives the relation

$$0 = (\lambda-m+2)\big[r^2\mathcal{M}_4(\bar g) - 2r[\partial_r^w, P_2(\bar g)]\big].$$

The proof is complete. □

Now we describe an alternative proof of the commutator relation in Theorem 7.1. The proof rests on a basic property of the *holographic Laplacian* $\mathcal{H}(h)(r)$ introduced in [Juhl 2013; 2016]. We recall that this operator is the Schrödinger-type operator

$$\mathcal{H}(h)(r) \stackrel{\text{def}}{=} -\delta_h(h_r^{-1}d) + \mathcal{U}(h)(r) \tag{7-1}$$

with the potential

$$\mathcal{U}(h)(r) \stackrel{\text{def}}{=} -w(r)^{-1}\big(\partial^2/\partial r^2 - (n-1)r^{-1}\partial/\partial r - \delta(h_r^{-1}d)\big)(w(r)). \tag{7-2}$$

The operator $\mathcal{H}(h)(r)$ should be viewed as a 1-parameter deformation of the Yamabe operator $\mathcal{H}(h)(0) = P_2(h)$. It is a key fact (see [Fefferman and Graham 2013; Juhl 2013]) that (the Taylor series of) this operator coincides with

$$\mathcal{G}(h)\Big(\frac{r^2}{4}\Big),$$

where

$$\mathcal{G}(h)(\rho) \stackrel{\text{def}}{=} \sum_{N\geq 1}\mathcal{M}_{2N}(h)\frac{\rho^{N-1}}{(N-1)!^2}$$

is a generating function of the so-called building block operators $\mathcal{M}_{2N}(h)$ of the GJMS operators of the metric h. In fact, any GJMS operator $P_{2N}(h)$ can be written as a linear combination

$$P_{2N}(h) = \sum_{|I|=N} n_I\mathcal{M}_{2I}(h), \quad n_I \in \mathbb{Z} \tag{7-3}$$

of compositions $\mathcal{M}_{2I} = \mathcal{M}_{2I_1}\cdots\mathcal{M}_{2I_r}$ for $I = (I_1, \ldots, I_r)$. For the details we refer to [Juhl 2013]. Now we consider the generating function $\mathcal{G}(\bar g(r))(\eta)$. It satisfies the basic relation

$$\mathcal{G}(\bar{g}(r))(\eta) = w(r)^{-1}\mathcal{G}(h)\Big(\frac{r^2}{4}+\eta\Big)w(r) + (\partial_r^w)^2 \tag{7-4}$$

of second-order operators acting on functions in (r, x). Note that the operator in the first term on the right-hand side only differentiates along M. The identity (7-4) follows by combining the relation between $\mathcal{G}(\bar{g})$ and the holographic Laplacian of $\bar{g}$ with an explicit formula [Juhl 2013, Theorem 7.2] for the Poincaré metric in normal form relative to $\bar{g}$ in terms of the Poincaré metric in normal form relative to h; for the details we refer to [Juhl 2012]. By expansion into powers of η, the relation (7-4) implies the identities

$$\mathcal{M}_{2N}(\bar{g}(r)) = \sum_{k\geq N} w(r)^{-1}\mathcal{M}_{2k}(h)w(r)\frac{(N-1)!}{(k-1)!(k-N)!}\Big(\frac{r^2}{4}\Big)^{k-N} \tag{7-5}$$

for $N \geq 2$. In turn, the latter relation yields the following commutator relations.

Theorem 7.2. *Let $N \geq 2$. Then*

$$r\mathcal{M}_{2N}(\bar{g}) = 2(N-1)[\partial_r^w, \mathcal{M}_{2N-2}(\bar{g})]. \tag{7-6}$$

Proof. Let $N \geq 3$. We use (7-5) and its relative

$$\mathcal{M}_{2N-2}(\bar{g}(r)) = \sum_{k\geq N-1} w^{-1}\mathcal{M}_{2k}(h)w\frac{(N-2)!}{(k-1)!(k-N+1)!}\Big(\frac{r^2}{4}\Big)^{k-N+1} \tag{7-7}$$

to calculate

$$\begin{aligned}
[\partial_r^w, \mathcal{M}_{2N-2}(\bar{g})] &= \partial_r^w\mathcal{M}_{2N-2}(\bar{g}) - \mathcal{M}_{2N-2}(\bar{g})\partial_r^w \\
&= w^{-1}\partial_r w\mathcal{M}_{2N-2}(\bar{g}) - \mathcal{M}_{2N-2}(\bar{g})w^{-1}\partial_r w \\
&= w^{-1}\partial_r\Big(\sum_{k\geq N-1}\mathcal{M}_{2k}(h)w\frac{(N-2)!}{(k-1)!(k-N+1)!}\Big(\frac{r^2}{4}\Big)^{k-1}\cdot\Big) \\
&\qquad - \sum_{k\geq N-1} w^{-1}\mathcal{M}_{2k}(h)\frac{(N-2)!}{(k-1)!(k-N+1)!}\Big(\frac{r^2}{4}\Big)^{k-1}\partial_r(w\,\cdot) \\
&= \frac{r}{2}\sum_{k\geq N-2} w^{-1}\mathcal{M}_{2k}(h)w\frac{(N-2)!}{(k-1)!(k-N)!}\Big(\frac{r^2}{4}\Big)^{k-2}\partial_r \\
&= \frac{r}{2(N-1)}\mathcal{M}_{2N}(\bar{g}).
\end{aligned}$$

This proves the assertion for $N \geq 3$. For $N = 2$, the identity (7-7) is to be replaced by

$$\mathcal{M}_2(\bar{g}) = \sum_{k\geq 1} w^{-1}\mathcal{M}_{2k}(h)w\frac{1}{(k-1)!(k-1)!}\Big(\frac{r^2}{4}\Big)^{k-1} + (\partial_r^w)^2.$$

Since the additional term $(\partial_r^w)^2$ commutes with ∂_r^w, the above arguments extend to that case. □

Theorem 7.1 is the special case $N = 2$ of Theorem 7.2.

Example 7.3. Let M be the sphere S^n with the round metric g_{S^n}. Then

$$\bar{g}(r) = dr^2 + (1 - r^2/4)^2 g_{S^n}.$$

But $\mathcal{M}_2(\bar{g}) = P_2(\bar{g})$ and

$$\mathcal{M}_{2N}(\bar{g}) = (N-1)!N!(1 - r^2/4)^{-N-1} P_2(g_{S^n})$$

for $N \geq 2$. These results can be derived from the identification of the holographic Laplacian $\mathcal{H}(\bar{g})$ with the generating series $\mathcal{G}(\bar{g})$ of the operators $\mathcal{M}_{2N}(\bar{g})$. For a direct proof see [Juhl 2013, Section 11.10]. Moreover, we have

$$\partial_r^w = \partial_r - \frac{n}{4} r (1 - r^2/4)^{-1}.$$

Hence we calculate

$$\begin{aligned} 2(N-1)[\partial_r^w, \mathcal{M}_{2N-2}(\bar{g})] &= 2(N-1)!(N-1)!\partial_r((1 - r^2/4)^{-N}) P_2(g_{S^n}) \\ &= N!(N-1)!(1 - r^2/4)^{-N-1} r P_2(g_{S^n}) \\ &= r \mathcal{M}_{2N}(\bar{g}) \end{aligned}$$

for $N \geq 3$. This proves (7-6) for $N \geq 3$. A direct calculation also confirms the case $N = 2$.

In turn, Theorem 7.2 leads to a simple compressed formula for the generating series $\mathcal{G}(\bar{g})$ and thus for the holographic Laplacian of $\bar{g}$. In order to formulate the result, we introduce the following notation. Let

$$R \circ \operatorname{ad}(\partial_r^w)(\cdot) \stackrel{\text{def}}{=} \frac{1}{r} \circ [\partial_r^w, \cdot].$$

Theorem 7.4. *Assume that (M^n, h) is real analytic of odd dimension n. Then*

$$\mathcal{H}(\bar{g})(\eta) = \mathcal{G}(\bar{g})\left(\frac{\eta^2}{4}\right) = \exp\left(R \circ \operatorname{ad}(\partial_r^w) \frac{\eta^2}{2}\right)(\mathcal{M}_2(\bar{g})). \tag{7-8}$$

Proof. By a repeated application of the identity (7-6) we obtain

$$\mathcal{M}_{2N}(\bar{g}) = 2^{N-1}(N-1)!(R \circ \operatorname{ad}(\partial_r^w))^{N-1}(\mathcal{M}_2(\bar{g}))$$

for $N \geq 2$. Using the natural convention $(R \circ \mathrm{ad}(\partial_r^w))^0 = \mathrm{Id}$, the latter identity also holds for $N = 1$. Hence

$$\begin{aligned}\mathcal{H}(\bar{g})(\eta) &= \sum_{N\geq 1} \mathcal{M}_{2N}(\bar{g}) \frac{1}{(N-1)!^2}\left(\frac{\eta^2}{4}\right)^{N-1} \\ &= \sum_{N\geq 1} (R \circ \mathrm{ad}(\partial_r^w))^{N-1}(\mathcal{M}_2(\bar{g})) \frac{1}{(N-1)!}\left(\frac{\eta^2}{2}\right)^{N-1} \\ &= \sum_{N\geq 0} (R \circ \mathrm{ad}(\partial_r^w))^{N}(\mathcal{M}_2(\bar{g})) \frac{1}{N!}\left(\frac{\eta^2}{2}\right)^{N} \\ &= \exp\left(R \circ \mathrm{ad}(\partial_r^w)\frac{\eta^2}{2}\right)(\mathcal{M}_2(\bar{g})).\end{aligned}$$

This completes the proof. □

Theorem 7.4 again clearly shows that $\mathcal{H}(\bar{g})(\eta)$ is a deformation of $\mathcal{M}_2(\bar{g}) = P_2(\bar{g})$.

We finish with brief comments on generalizations of the present theory to differential forms. The theory of differential symmetry breaking operators on functions has a natural extension to differential forms [Fischmann et al. 2016; Kobayashi et al. 2016]. Curved versions of that theory deal with residue families acting on differential forms. Their theory will be developed elsewhere. Residue families on differential forms are expected to have analogous descriptions in terms of compositions of shift operators on differential forms. Results in [Fischmann et al. 2019] confirm that picture in the flat case. Curved analogs of the degenerate Laplacian acting on differential forms were studied in [Gover et al. 2015] in terms of tractor calculus.

Acknowledgment

Juhl is grateful to Aarhus University for hospitality, stimulating atmosphere and financial support.

References

[Baum and Juhl 2010] H. Baum and A. Juhl, *Conformal differential geometry: Q-curvature and conformal holonomy*, Oberwolfach Seminars **40**, Birkhäuser, Basel, 2010. MR Zbl

[Branson 1995] T. P. Branson, “Sharp inequalities, the functional determinant, and the complementary series”, *Trans. Amer. Math. Soc.* **347**:10 (1995), 3671–3742. MR Zbl

[Clerc 2017] J.-L. Clerc, “Another approach to Juhl’s conformally covariant differential operators from S^n to S^{n-1}”, *SIGMA Symmetry Integrability Geom. Methods Appl.* **13** (2017), art. no. 026, 18 pp. MR Zbl

[Fefferman and Graham 2012] C. Fefferman and C. R. Graham, *The ambient metric*, Annals of Mathematics Studies **178**, Princeton University Press, 2012. MR Zbl

[Fefferman and Graham 2013] C. Fefferman and C. R. Graham, “Juhl’s formulae for GJMS operators and Q-curvatures”, *J. Amer. Math. Soc.* **26**:4 (2013), 1191–1207. MR Zbl

[Fefferman and Hirachi 2003] C. Fefferman and K. Hirachi, "Ambient metric construction of Q-curvature in conformal and CR geometries", *Math. Res. Lett.* **10**:5-6 (2003), 819–831. MR Zbl

[Fischmann et al. 2016] M. Fischmann, A. Juhl, and P. Somberg, "Conformal symmetry breaking differential operators on differential forms", preprint, 2016. To appear in *Mem. Amer. Math. Soc.* arXiv

[Fischmann et al. 2019] M. Fischmann, B. Ørsted, and P. Somberg, "Bernstein–Sato identities and conformal symmetry breaking operators", *J. Funct. Anal.* **277**:11 (2019), 108219, 36. MR Zbl

[Frahm and Ørsted 2019] J. Frahm and B. Ørsted, "Knapp–Stein type intertwining operators for symmetric pairs, II: The translation principle and intertwining operators for spinors", *SIGMA Symmetry Integrability Geom. Methods Appl.* **15** (2019), art. no. 084, 50 pp. MR Zbl

[Gover 2010] A. R. Gover, "Almost Einstein and Poincaré–Einstein manifolds in Riemannian signature", *J. Geom. Phys.* **60**:2 (2010), 182–204. MR Zbl

[Gover and Hirachi 2004] A. R. Gover and K. Hirachi, "Conformally invariant powers of the Laplacian—a complete nonexistence theorem", *J. Amer. Math. Soc.* **17**:2 (2004), 389–405. MR Zbl

[Gover and Peterson 2018] A. R. Gover and L. J. Peterson, "Conformal boundary operators, T-curvatures, and conformal fractional Laplacians of odd order", preprint, 2018. arXiv

[Gover and Waldron 2014] A. R. Gover and A. Waldron, "Boundary calculus for conformally compact manifolds", *Indiana Univ. Math. J.* **63**:1 (2014), 119–163. MR Zbl

[Gover et al. 2015] A. R. Gover, E. Latini, and A. Waldron, *Poincaré–Einstein holography for forms via conformal geometry in the bulk*, Mem. Amer. Math. Soc. **1106**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Graham 1992] C. R. Graham, "Conformally invariant powers of the Laplacian, II: Nonexistence", *J. London Math. Soc.* (2) **46**:3 (1992), 566–576. MR Zbl

[Graham 2000] C. R. Graham, "Volume and area renormalizations for conformally compact Einstein metrics", *Rend. Circ. Mat. Palermo* (2) *Suppl.* **63** (2000), 31–42. MR Zbl

[Graham and Juhl 2007] C. R. Graham and A. Juhl, "Holographic formula for Q-curvature", *Adv. Math.* **216**:2 (2007), 841–853. MR Zbl

[Graham and Lee 1991] C. R. Graham and J. M. Lee, "Einstein metrics with prescribed conformal infinity on the ball", *Adv. Math.* **87**:2 (1991), 186–225. MR Zbl

[Graham and Zworski 2003] C. R. Graham and M. Zworski, "Scattering matrix in conformal geometry", *Invent. Math.* **152**:1 (2003), 89–118. MR Zbl

[Graham et al. 1992] C. R. Graham, R. Jenne, L. J. Mason, and G. A. J. Sparling, "Conformally invariant powers of the Laplacian, I: Existence", *J. London Math. Soc.* (2) **46**:3 (1992), 557–565. MR Zbl

[Helgason 1984] S. Helgason, *Groups and geometric analysis: integral geometry, invariant differential operators, and spherical functions*, Pure and Applied Mathematics **113**, Academic Press, Orlando, FL, 1984. MR Zbl

[Juhl 2009] A. Juhl, *Families of conformally covariant differential operators, Q-curvature and holography*, Progress in Mathematics **275**, Birkhäuser, Basel, 2009. MR Zbl

[Juhl 2011] A. Juhl, "Holographic formula for Q-curvature, II", *Adv. Math.* **226**:4 (2011), 3409–3425. MR Zbl

[Juhl 2012] A. Juhl, "On the building block operators of the GJMS operators", unpublished notes, 2012. BIRS workshop.

[Juhl 2013] A. Juhl, "Explicit formulas for GJMS-operators and Q-curvatures", *Geom. Funct. Anal.* **23**:4 (2013), 1278–1370. MR Zbl

[Juhl 2014] A. Juhl, "On the recursive structure of Branson's Q-curvatures", *Math. Res. Lett.* **21**:3 (2014), 495–507. MR Zbl

[Juhl 2016] A. Juhl, "Heat kernel expansions, ambient metrics and conformal invariants", *Adv. Math.* **286** (2016), 545–682. MR Zbl

[Kobayashi and Pevzner 2016] T. Kobayashi and M. Pevzner, "Differential symmetry breaking operators, I: General theory and F-method", *Selecta Math.* (*N.S.*) **22**:2 (2016), 801–845. MR Zbl

[Kobayashi and Speh 2015] T. Kobayashi and B. Speh, *Symmetry breaking for representations of rank one orthogonal groups*, Mem. Amer. Math. Soc. **1126**, Amer. Math. Soc., Providence, RI, 2015. MR Zbl

[Kobayashi and Speh 2018] T. Kobayashi and B. Speh, *Symmetry breaking for representations of rank one orthogonal groups II*, Lecture Notes in Mathematics **2234**, Springer, 2018. MR Zbl

[Kobayashi et al. 2015] T. Kobayashi, B. Ørsted, P. Somberg, and V. Souček, "Branching laws for Verma modules and applications in parabolic geometry, I", *Adv. Math.* **285** (2015), 1796–1852. MR Zbl

[Kobayashi et al. 2016] T. Kobayashi, T. Kubo, and M. Pevzner, *Conformal symmetry breaking operators for differential forms on spheres*, Lecture Notes in Mathematics **2170**, Springer, 2016. MR Zbl

[Patterson and Perry 2001] S. J. Patterson and P. A. Perry, "The divisor of Selberg's zeta function for Kleinian groups", *Duke Math. J.* **106**:2 (2001), 321–390. MR Zbl

Received August 20, 2018.

ANDREAS JUHL
INSTITUT FÜR MATHEMATIK
HUMBOLT-UNIVERSITÄT
BERLIN
GERMANY
juhl.andreas@googlemail.com

BENT ØRSTED
DEPARTMENT OF MATHEMATICS
AARHUS UNIVERSITY
DENMARK
orsted@math.au.dk

PACIFIC JOURNAL OF MATHEMATICS
Vol. 308, No. 1, 2020
https://doi.org/10.2140/pjm.2020.308.161

DIFFERENTIAL-HENSELIANITY AND MAXIMALITY OF ASYMPTOTIC VALUED DIFFERENTIAL FIELDS

NIGEL PYNN-COATES

We show that asymptotic (valued differential) fields have unique maximal immediate extensions. Connecting this to differential-henselianity, we prove that any differential-henselian asymptotic field is differential-algebraically maximal, removing the assumption of monotonicity from a theorem of Aschenbrenner, van den Dries, and van der Hoeven. Finally, we use this result to show the existence and uniqueness of differential-henselizations of asymptotic fields.

1. Introduction

The basic objects of this paper are valued differential fields (assumed throughout to have equicharacteristic 0) with small derivation and their *extensions*, by which we mean valued differential field extensions with small derivation. By Zorn's lemma, any valued differential field K with small derivation has an immediate extension that is *maximal* in the sense that it has no proper immediate extension. By the main result of [ADH 2018], such extensions are spherically complete, and thus maximal as valued fields in the usual sense. Hence by [Kaplansky 1942] any two maximal immediate extensions of K are isomorphic as valued fields over K. If K is additionally assumed to be monotone and to have linearly surjective differential residue field, then Aschenbrenner, van den Dries, and van der Hoeven showed in [ADH 2017, Theorem 7.4.3] that they are isomorphic as valued *differential* fields over K and conjectured in [ADH 2018] that monotonicity could be removed from this result. Van den Dries and Pynn-Coates [2019] showed that this conjecture holds when the value group is the union of its convex subgroups of finite archimedean rank. Here, we prove the conjecture when the valued differential field is asymptotic, a condition opposite in spirit to monotonicity.

Theorem 3.5. *If an asymptotic valued differential field K has small derivation and linearly surjective differential residue field, then any two maximal immediate extensions of K are isomorphic over K.*

MSC2010: 12H05, 12J10.
Keywords: valued differential fields, asymptotic fields, immediate extensions, differential-henselianity, differential Newton diagrams.

We also show the uniqueness of differentially algebraic immediate extensions of K that are differential-algebraically maximal. This was likewise proved earlier in the monotone setting [ADH 2017, Theorem 7.4.3], with the case of many constants essentially going back to [Scanlon 2000].

Theorem 3.5 has a consequence related to the topics discussed in Matusinski's survey [2014] transposed to this setting. Let K be an asymptotic valued differential field with small derivation. By [ADH 2018], K has an immediate extension L that is asymptotic and spherically complete, so L is isomorphic as a valued differential field to a Hahn field (also called a "generalized series field" in [Matusinski 2014]) endowed with a derivation making it an asymptotic valued differential field with small derivation. Hence if K has linearly surjective differential residue field, then any other immediate differential Hahn field extension of K that is asymptotic is also isomorphic to L over K by Theorem 3.5.

Since there is an equivalence between algebraic maximality and henselianity for valued fields of equicharacteristic 0, one might hope for a similar relationship in the setting of valued differential fields after adding the prefix "differential", but it turns out to depend on the interaction between the valuation and the derivation. In fact, any differential-algebraically maximal valued differential field with small derivation and linearly surjective differential residue field is differential-henselian [ADH 2017, Theorem 7.0.1], but the converse fails, even in the setting of monotone fields with many constants [ADH 2017, example after Corollary 7.4.5]. In contrast, we show that the converse holds in the case of few constants, removing entirely the monotonicity assumption in [ADH 2017, Theorem 7.0.3]:

Theorem 3.6. *If K is a valued differential field with small derivation and few constants that is differential-henselian, then K is differential-algebraically maximal.*

Note that by [ADH 2017, Lemmas 7.1.8 and 9.1.1], any K as in the theorem above is in fact asymptotic. Using Theorem 3.6 we then prove the following generalization of [van den Dries and Pynn-Coates 2019, Theorem 1.3].

Theorem 3.7. *If K is an asymptotic valued differential field with small derivation and linearly surjective differential residue field, then K has a differential-henselization which is minimal, and so any two differential-henselizations of K are isomorphic over K.*

1.A. ***Basic definitions and notation.*** To understand these results, we define the necessary conditions after setting up basic notation, which we keep close to that of [ADH 2017]. We let d, i, m, n, and r range over $\mathbb{N} = \{0, 1, 2, \dots\}$. Throughout, K is a *valued differential field*, that is, a field of characteristic 0 together with a surjective map $v : K^\times \to \Gamma$, where Γ is a (totally) ordered abelian group called the *value group* of K, and a map $\partial : K \to K$ satisfying $v(\mathbb{Q}^\times) = \{0\}$ and, for all x, y in their domain, (V1) $v(xy) = v(x) + v(y)$;

(V2) $v(x + y) \geqslant \min\{v(x), v(y)\}$ whenever $x + y \neq 0$;

(D1) $\partial(x + y) = \partial(x) + \partial(y)$;

(D2) $\partial(xy) = x\partial(y) + \partial(x)y$.

We introduce a symbol $\infty \notin \Gamma$ and extend the ordering to $\Gamma \cup \{\infty\}$ by $\infty > \Gamma$. We also set $\infty + \gamma = \gamma + \infty = \infty + \infty := \infty$ for all $\gamma \in \Gamma$, allowing us to extend v to K by setting $v(0) := \infty$. We also set $\Gamma^{\neq} := \Gamma \setminus \{0\}$. We let $\mathcal{O} := \{a \in K : v(a) \geqslant 0\}$ be the *valuation ring* of K and $\mathcal{o} := \{a \in K : v(a) > 0\}$ its (unique) maximal ideal. We also set $\mathcal{O}^{\neq} := \mathcal{O} \setminus \{0\}$ and $\mathcal{o}^{\neq} := \mathcal{o} \setminus \{0\}$. Then $\boldsymbol{k} := \mathcal{O}/\mathcal{o}$ denotes the *residue field* of K. As it is often more intuitive, we define for $a, b \in K$

$$a \preccurlyeq b \;\Leftrightarrow\; v(a) \geqslant v(b), \qquad a \prec b \;\Leftrightarrow\; v(a) > v(b),$$
$$a \asymp b \;\Leftrightarrow\; v(a) = v(b), \qquad a \sim b \;\Leftrightarrow\; a - b \prec b.$$

Regarding the derivation, we usually write a' for $\partial(a)$ if the derivation is clear from context, and the *field of constants* of K is denoted by $C := \{a \in K : \partial(a) = 0\}$. For another valued differential field L, we apply the subscript L to these symbols; for example, $\mathcal{O}_L$ denotes the valuation ring of L. Recall that a valued field extension L of K is said to be *immediate* if $\boldsymbol{k}_L = \boldsymbol{k}$ and $\Gamma_L = \Gamma$, where we identify $\boldsymbol{k}$ with a subfield of $\boldsymbol{k}_L$ and Γ with an ordered subgroup of Γ_L in the usual way. We call a pseudocauchy sequence a *pc-sequence* and refer the reader to [ADH 2017, §2.2 and §3.2] for definitions and basic facts about them.

We let $K\{Y\} := K[Y, Y', Y'', \dots]$ denote the ring of differential polynomials over K (we hereinafter shorten “differential-polynomial” to “d-polynomial”) and set $K\{Y\}^{\neq} := K\{Y\} \setminus \{0\}$. Let P range over $K\{Y\}^{\neq}$. The *order* of P is the smallest r such that $P \in K[Y, Y', \dots, Y^{(r)}]$. Its *degree* is its total degree. If r is the order of P, m its degree in $Y^{(r)}$, and n its (total) degree, then the *complexity* of P is the triple $c(P) := (r, m, n)$; complexities are ordered lexicographically. For $\boldsymbol{i} = (i_0, \dots, i_r) \in \mathbb{N}^{1+r}$, we set $Y^{\boldsymbol{i}} := Y^{i_0}(Y')^{i_1} \cdots (Y^{(r)})^{i_r}$. If P has order at most r, then we decompose P as $\sum_{\boldsymbol{i}} P_{\boldsymbol{i}} Y^{\boldsymbol{i}}$, where $\boldsymbol{i}$ ranges over $\mathbb{N}^{1+r}$. We also sometimes decompose P into its homogeneous parts, so let P_d denote the homogeneous part of P of degree d and set $P_{\leqslant d} := \sum_{i \leqslant d} P_i$ and $P_{>d} := \sum_{i>d} P_i$. Letting $|\boldsymbol{i}| := i_0 + \cdots + i_r$, we note that $P_d = \sum_{|\boldsymbol{i}|=d} P_{\boldsymbol{i}} Y^{\boldsymbol{i}}$. The *multiplicity of P at* 0, denoted by $\operatorname{mul} P$, is the least d with $P_d \neq 0$. We often use, for $a \in K$, the additive and multiplicative conjugates of P by a defined by $P_{+a}(Y) := P(a + Y)$ and $P_{\times a}(Y) := P(aY)$. For convenience, we also write P_{-a} for $P_{+(-a)}$. Note that $(P_{+a})_{+b} = (P_{+b})_{+a} = P_{+(a+b)}$ for $b \in K$, which we write P_{+a+b}. We define P_{+a-b} likewise. The *multiplicity of P at a* is $\operatorname{mul} P_{+a}$. Note that $(P_d)_{\times a} = (P_{\times a})_d$, which we denote by $P_{d,\times a}$. For more on such conjugation, see [ADH 2017, §4.3]. We extend the derivation of K to $K\{Y\}$ in the natural way, and we also extend v to $K\{Y\}$

by setting $v(P)$ to be the minimum valuation of the coefficients of P. The relations $\preccurlyeq$, $\prec$, $\asymp$, and $\sim$ are also extended to $K\{Y\}$ in the obvious way. We recall how $v(P)$ behaves under additive and multiplicative conjugation of P in Section 2.E.

There are two conditions we sometimes impose connecting the valuation and the derivation. First, we say K is *asymptotic* if, for all $f, g \in \mathcal{o}^{\neq}$,

$$f \prec g \iff f' \prec g'.$$

If K is asymptotic, then ∂ is continuous with respect to the valuation topology on K [ADH 2017, Corollary 9.1.5]. It follows immediately from the definition that if K is asymptotic, then $v(C^{\times}) = \{0\}$, in which case we say that K has *few constants*. This weaker condition is assumed in some lemmas, so as to delay the assumption that K is asymptotic until Section 6. The class of asymptotic fields includes for example the ordered (valued) differential field of logarithmic-exponential transseries studied in [ADH 2017], and, more generally, the class of differential-valued fields introduced in [Rosenlicht 1980].

The second condition is more basic and is assumed throughout: We say that K has *small derivation* if $\partial\mathcal{o} \subseteq \mathcal{o}$. Small derivation also implies that ∂ is continuous with respect to the valuation topology on K; in fact, ∂ is continuous if and only if $\partial\mathcal{o} \subseteq a\mathcal{o}$ for some $a \in K^{\times}$ [ADH 2017, Lemma 4.4.7]. It also implies that $\partial\mathcal{O} \subseteq \mathcal{O}$ [ADH 2017, Lemma 4.4.2], so ∂ induces a derivation on $\boldsymbol{k}$. Then we say that $\boldsymbol{k}$ is *r-linearly surjective* if, for all $a_0, \dots, a_r \in \boldsymbol{k}$ with $a_i \neq 0$ for some $i \leqslant r$, the equation $1 + a_0 y + a_1 y' + \cdots + a_r y^{(r)} = 0$ has a solution in $\boldsymbol{k}$. We call $\boldsymbol{k}$ *linearly surjective* if $\boldsymbol{k}$ is r-linearly surjective for each r. Generalizing the notion of henselianity for valued fields, we say that K is *r-differential-henselian* (*r-d-henselian* for short) if

(rDH1) $\boldsymbol{k}$ is r-linearly surjective, and

(rDH2) whenever $P \in \mathcal{O}\{Y\}$ has order at most r and satisfies $P_0 \prec 1$ and $P_1 \asymp 1$, there is $y \in \mathcal{o}$ with $P(y) = 0$.

The following equivalent formulation is often used without comment. Its proof uses the equalizer theorem (see Theorem 4.1), a key result from [ADH 2017] which also underlies the results of this paper.

Lemma 1.1 [ADH 2017, Lemmas 7.1.1 and 7.2.1]. *We have that K is r-d-henselian if and only if for every $P \in \mathcal{O}\{Y\}$ of order at most r satisfying $P_1 \asymp 1$ and $P_i \prec 1$ for all $i \geqslant 2$, there is $y \in \mathcal{O}$ with $P(y) = 0$.*

We say that K is *differential-henselian* (*d-henselian* for short) if it is r-d-henselian for each r. These definitions are due to Aschenbrenner, van den Dries, and van der Hoeven [ADH 2017, Chapter 7], although earlier notions were considered by Scanlon [2000] for monotone fields and F.-V. Kuhlmann [2011] for differential-valued fields. Connecting this to asymptoticity, we note that if K is 1-d-henselian and has few constants, then it is asymptotic [ADH 2017, Lemmas 7.1.8 and 9.1.1].

Throughout, by an *extension of* K we mean a valued differential field extension of K with small derivation; similarly, *embedding* means "valued differential field embedding". In analogy with the notion of a henselization of a valued field, we defined in [van den Dries and Pynn-Coates 2019] the notion of a differential-henselization of a valued differential field: we call an extension L of K a *differential-henselization* (*d-henselization* for short) of K if it is an immediate d-henselian extension of K that embeds over K into any immediate d-henselian extension of K. Finally, we call K *maximal* if it has no proper immediate extension and *differential-algebraically maximal* (*d-algebraically maximal* for short) if it has no proper differentially algebraic immediate extension. Recall from [ADH 2017, Chapter 7] that if the derivation induced on $\boldsymbol{k}$ is nontrivial, then K is d-algebraically maximal if and only if every pc-sequence in K of d-algebraic type over K has a pseudolimit in K (see Section 2.C for this notion). It is also worth pointing out that by the main result of [ADH 2018], K is maximal in this sense if and only if K is maximal as a valued field in the usual sense, which in turn is equivalent to every pc-sequence in K having a pseudolimit in K.

1.B. ***Outline.*** The main technical tool of the paper is Proposition 3.1, and assuming this we prove Theorems 3.5, 3.6, and 3.7 in Section 3. We also use Proposition 3.1 to obtain analogues of these results relativized to d-polynomials of a fixed order in Section 3.C. The rest of the paper after Section 3 is devoted to proving Proposition 3.1, and the strategy closely follows the approach taken to prove [ADH 2017, Proposition 14.5.1], which is an analogue of Proposition 3.1 in the setting of ω-free H-asymptotic differential-valued fields.

First, we adapt the differential newton diagram method of [ADH 2017, §13.5] to the setting of valued differential fields with small derivation and divisible value group in Section 4, which relies in an essential way on the equalizer theorem [ADH 2017, Theorem 6.0.1]; this is where divisibility is used. The main results are Proposition 4.4 and Corollary 4.5, which are then connected to pc-sequences in Section 4.A.

From there, we proceed to study asymptotic differential equations in Section 5, with the main technical notion being that of an unraveller, adapted from [ADH 2017, §13.8]. There are three key steps in this section. First, we establish the existence of an unraveller that is a pseudolimit of a pseudocauchy sequence in Lemma 5.8, via Proposition 5.5. Second, we reduce the degree of an asymptotic differential equation in Lemma 5.10. Third, we find a solution of an asymptotic differential equation in a d-henselian field that approximates an element in an extension of that field in Section 5.C.

The penultimate section, Section 6, based on [ADH 2017, §14.4], is quite technical. It combines many results from the previous sections and culminates in

Proposition 6.1 and Corollary 6.14, which is essential to the proof of Proposition 3.1. One of the main steps here is Lemma 6.11, which allows us to use Lemma 5.10 to reduce the degree of an asymptotic differential equation.

There are four salient differences from the approach in [ADH 2017]. The first is that the "dominant part" and "dominant degree" of d-polynomials replace their more technical cousins "newton polynomial" and "newton degree", leading to the simplification of some proofs. The second is that since K is not assumed to be H-asymptotic, we replace the convex valuation on Γ given by $v(g) \mapsto v(g'/g)$, for $g \in K^\times$ with $g \not\asymp 1$, with that given by considering archimedean classes. Third, under the assumption of ω-freeness, newton polynomials are in $C[Y](Y')^n$ for some n, but dominant parts need not have this special form. This leads to changes in Section 6, such as the need to take partial derivatives with respect to higher order derivatives of Y than just Y'. Finally, Lemma 3.2 enables the generalization of Proposition 3.1 to nondivisible value group in Proposition 3.3. (An analogous lemma was used in [Pynn-Coates 2019] to remove the assumption of divisible value group from the corresponding results in [ADH 2017, Chapter 14].)

1.C. ***Review of assumptions.*** Throughout, K is a valued differential field with nontrivial (surjective) valuation $v : K^\times \to \Gamma$ and nontrivial derivation $\partial : K \to K$. The valuation ring is $\mathcal{O}$ with maximal ideal $\circ$, and we further assume that K has small derivation, i.e., $\partial\circ \subseteq \circ$. Then the differential residue field is $\boldsymbol{k} = \mathcal{O}/\circ$. The field of constants is denoted by C. Additional assumptions on K, Γ, and $\boldsymbol{k}$ are indicated as needed. "Extension" is short for "valued differential field extension with small derivation" and "embedding" is short for "valued differential field embedding". We let d, i, m, n, and r range over $\mathbb{N} = \{0, 1, 2, \ldots\}$.

2. Preliminaries

Throughout this section, $P \in K\{Y\}^{\neq}$.

2.A. ***Dominant parts of differential polynomials.*** We present in this subsection the notion of the dominant part of a d-polynomial over K when K has a monomial group $\mathfrak{M}$, i.e., a subgroup of $K^\times$ that is mapped bijectively onto Γ by v. This assumption yields slightly improved versions of lemmas from [ADH 2017, §6.6], where this notion is developed without the monomial group assumption. All the statements about dominant degree and dominant multiplicity given here remain true in that greater generality, and we freely use them later even when K may not have a monomial group. The proofs are essentially the same, so are omitted.

Assumption. In this subsection, K has a monomial group $\mathfrak{M}$.

We let $\mathfrak{d}_P \in \mathfrak{M}$ be the unique monomial such that $\mathfrak{d}_P \asymp P$. For $Q = 0 \in K\{Y\}$, we set $\mathfrak{d}_Q := 0$.

Definition. Since $\mathfrak{d}_P^{-1}P \in \mathcal{O}\{Y\}$, we define the *dominant part* of P to be the d-polynomial

$$D_P := \overline{\mathfrak{d}_P^{-1}P} = \sum_{\boldsymbol{i}} (\overline{P_{\boldsymbol{i}}/\mathfrak{d}_P})Y^{\boldsymbol{i}} \in \boldsymbol{k}\{Y\}^{\neq}.$$

For $Q = 0 \in K\{Y\}$, we set $D_Q := 0 \in \boldsymbol{k}\{Y\}$.

Then $\deg D_P \leqslant \deg P$ and $\operatorname{ord} D_P \leqslant \operatorname{ord} P$. We call $\operatorname{ddeg} P := \deg D_P$ the *dominant degree* of P and $\operatorname{dmul} P := \operatorname{mul} D_P$ the *dominant multiplicity of P at* 0.

Note that if P is homogeneous of degree d, then so is D_P.

Lemma 2.1. *Let $Q \in K\{Y\}$. Then*

(i) *if $P \succ Q$, then $D_{P+Q} = D_P$;*

(ii) *if $P \asymp Q$ and $P + Q \asymp P$, then $D_{P+Q} = D_P + D_Q$;*

(iii) $D_{PQ} = D_P D_Q$.

Proof. Part (ii) is not in [ADH 2017], so we give a proof. Suppose $P \asymp Q$ and $P + Q \asymp P$, so $\mathfrak{d}_{P+Q} = \mathfrak{d}_P = \mathfrak{d}_Q$. Then

$$D_{P+Q} = \sum_{\boldsymbol{i}} \overline{(P+Q)_{\boldsymbol{i}}/\mathfrak{d}_{P+Q}}\, Y^{\boldsymbol{i}} = \sum_{\boldsymbol{i}} (\overline{P_{\boldsymbol{i}}/\mathfrak{d}_P} + \overline{Q_{\boldsymbol{i}}/\mathfrak{d}_Q})Y^{\boldsymbol{i}} = D_P + D_Q. \qquad \square$$

Lemma 2.2. *Let $a \in K$ with $a \preccurlyeq 1$. Then*

(i) $D_{P_{+a}} = (D_P)_{+\bar{a}}$, *and thus* $\operatorname{ddeg} P_{+a} = \operatorname{ddeg} P$;

(ii) *if $a \asymp 1$, then* $D_{P_{\times a}} = (D_P)_{\times \bar{a}}$, $\operatorname{dmul} P_{\times a} = \operatorname{dmul} P$, *and* $\operatorname{ddeg} P_{\times a} = \operatorname{ddeg} P$;

(iii) *if $a \prec 1$, then* $\operatorname{ddeg} P_{\times a} \leqslant \operatorname{dmul} P$.

It follows from (ii) that $\operatorname{ddeg} P_{\times g}$ and $\operatorname{dmul} P_{\times g}$ only depend on vg, for $g \in K^{\times}$. The next five results are exactly as in [ADH 2017, §6.6], but are recalled here for the reader's convenience.

Corollary 2.3 [ADH 2017, Corollary 6.6.6]. *If $f, g \in K$ and $h \in K^{\times}$ satisfy $f - g \preccurlyeq h$, then*

$$\operatorname{ddeg} P_{+f,\times h} = \operatorname{ddeg} P_{+g,\times h}.$$

Corollary 2.4 [ADH 2017, Corollary 6.6.7]. *Let $f, g \in K^{\times}$. Then* $\operatorname{mul} P = \operatorname{mul}(P_{\times f}) \leqslant \operatorname{ddeg} P_{\times f}$ *and*

$$f \prec g \implies \operatorname{dmul} P_{\times f} \leqslant \operatorname{ddeg} P_{\times f} \leqslant \operatorname{dmul} P_{\times g} \leqslant \operatorname{ddeg} P_{\times g}.$$

Below, we let $\mathcal{E} \subseteq K^{\times}$ be nonempty and such that for $f, g \in K^{\times}$, $f \preccurlyeq g \in \mathcal{E}$ implies $f \in \mathcal{E}$. In this case, we say that $\mathcal{E}$ is *$\preccurlyeq$-closed*, and we consider the *dominant degree of P on $\mathcal{E}$* defined by

$$\operatorname{ddeg}_{\mathcal{E}} P := \max\{\operatorname{ddeg} P_{\times f} : f \in \mathcal{E}\}.$$

Note that $\preccurlyeq$-closed sets correspond to nonempty upward-closed subsets of Γ. If for $\gamma \in \Gamma$ we have $\mathcal{E} = \{f \in K^\times : vf \geqslant \gamma\}$, then $\operatorname{ddeg}_{\geqslant\gamma} P := \operatorname{ddeg}_{\mathcal{E}} P$. For any $g \in K^\times$ with $vg = \gamma$ we set $\operatorname{ddeg}_{\preccurlyeq g} P := \operatorname{ddeg}_{\geqslant\gamma} P$, and by the previous result we have $\operatorname{ddeg}_{\preccurlyeq g} P = \operatorname{ddeg} P_{\times g}$. We define $\operatorname{ddeg}_{>\gamma} P$ and $\operatorname{ddeg}_{\prec g} P$ analogously.

Lemma 2.5 [ADH 2017, Lemma 6.6.9]. *If $v(\mathcal{E})$ has no smallest element, then*

$$\operatorname{ddeg}_{\mathcal{E}} P = \max\{\operatorname{dmul}(P_{\times f}) : f \in \mathcal{E}\}.$$

Lemma 2.6 [ADH 2017, Lemma 6.6.10]. *If $f \in \mathcal{E}$, then* $\operatorname{ddeg}_{\mathcal{E}} P_{+f} = \operatorname{ddeg}_{\mathcal{E}} P$.

Corollary 2.7 [ADH 2017, Corollary 6.6.11]. *Suppose that* $\operatorname{ddeg}_{\mathcal{E}} P = 1$ *and $y \in \mathcal{E}$ satisfies $P(y) = 0$. If $f \in \mathcal{E}$, then*

$$\operatorname{mul} P_{+y,\times f} = \operatorname{dmul} P_{+y,\times f} = \operatorname{ddeg} P_{+y,\times f} = 1.$$

2.B. ***Dominant degree in a cut.*** We recall the notion of "dominant degree in a cut" from [van den Dries and Pynn-Coates 2019] and some basic properties proved there. First, for later use we mention that the condition $\operatorname{ddeg} P \geqslant 1$ is necessary for the existence of a zero $f \preccurlyeq 1$ of P in an extension of K:

Lemma 2.8 [van den Dries and Pynn-Coates 2019, Lemma 2.1]. *Let $g \in K^\times$ and suppose that $P(f) = 0$ for some $f \preccurlyeq g$ in some extension of K. Then* $\operatorname{ddeg} P_{\times g} \geqslant 1$.

In the rest of this section, (a_ρ) is a pc-sequence in K with $\gamma_\rho := v(a_{\rho+1} - a_\rho)$; here and later $\rho + 1$ denotes the immediate successor of ρ in the well-ordered set of indices.

Lemma 2.9 [van den Dries and Pynn-Coates 2019, Lemma 2.2]. *There is an index ρ_0 and a number $d(P, (a_\rho)) \in \mathbb{N}$ such that for all $\rho > \rho_0$,*

$$\operatorname{ddeg}_{\geqslant\gamma_\rho} P_{+a_\rho} = d(P, (a_\rho)).$$

If (b_σ) is a pc-sequence in K equivalent to (a_ρ), we have $d(P, (a_\rho)) = d(P, (b_\sigma))$.

We associate to each pc-sequence (a_ρ) in K its *cut in K*, denoted by $c_K(a_\rho)$, such that if (b_σ) is a pc-sequence in K, then

$$c_K(a_\rho) = c_K(b_\sigma) \iff (b_\sigma) \text{ is equivalent to } (a_\rho).$$

Throughout, $\boldsymbol{a} := c_K(a_\rho)$ and if L is an extension of K, then $\boldsymbol{a}_L := c_L(a_\rho)$. Note that $c_K(a_\rho + y)$ for $y \in K$ depends only on $\boldsymbol{a}$ and y, so we let $\boldsymbol{a} + y$ denote $c_K(a_\rho + y)$ and $\boldsymbol{a} - y$ denote $c_K(a_\rho - y)$. Similarly, $c_K(a_\rho y)$ for $y \in K^\times$ depends only on $\boldsymbol{a}$ and y, so we let $\boldsymbol{a} \cdot y$ denote $c_K(a_\rho y)$.

Definition. The *dominant degree of P in the cut of (a_ρ)*, denoted by $\operatorname{ddeg}_{\boldsymbol{a}} P$, is the natural number $d(P, (a_\rho))$ from the previous lemma.

Lemma 2.10 [van den Dries and Pynn-Coates 2019, Lemma 2.3]. *Dominant degree in a cut has the following properties*:

(i) $\operatorname{ddeg}_{\boldsymbol{a}} P \leqslant \deg P$;

(ii) $\operatorname{ddeg}_{\boldsymbol{a}} P_{+y} = \operatorname{ddeg}_{\boldsymbol{a}+y} P$ *for* $y \in K$;

(iii) *if* $y \in K$ *and* vy *is in the width of* (a_ρ), *then* $\operatorname{ddeg}_{\boldsymbol{a}} P_{+y} = \operatorname{ddeg}_{\boldsymbol{a}} P$;

(iv) $\operatorname{ddeg}_{\boldsymbol{a}} P_{\times y} = \operatorname{ddeg}_{\boldsymbol{a}\cdot y} P$ *for* $y \in K^\times$;

(v) *if* $Q \in K\{Y\}^{\neq}$, *then* $\operatorname{ddeg}_{\boldsymbol{a}} PQ = \operatorname{ddeg}_{\boldsymbol{a}} P + \operatorname{ddeg}_{\boldsymbol{a}} Q$;

(vi) *if* $P(\ell) = 0$ *for some pseudolimit* ℓ *of* (a_ρ) *in an extension of* K, *then* $\operatorname{ddeg}_{\boldsymbol{a}} P \geqslant 1$;

(vii) *if* L *is an extension of* K, *then* $\operatorname{ddeg}_{\boldsymbol{a}} P = \operatorname{ddeg}_{\boldsymbol{a}_L} P$.

2.C. ***Constructing immediate extensions.*** We review how to construct immediate extensions by evaluating d-polynomials along pc-sequences. Recall from [ADH 2017, §4.4] that a pc-sequence (a_ρ) in K is of *d-algebraic type over* K if there is an equivalent pc-sequence (b_σ) in K and a $P \in K\{Y\}$ such that $P(b_\sigma) \rightsquigarrow 0$; such a P of minimal complexity is called a *minimal d-polynomial of* (a_ρ) *over* K. If no such (b_σ) and P exist, then (a_ρ) is of *d-transcendental type over* K.

Assumption. In this subsection, the derivation induced on $\boldsymbol{k}$ is nontrivial.

Lemma 2.11 [ADH 2017, Lemma 6.8.1]. *Let* (a_ρ) *be a pc-sequence in* K *with pseudolimit* $a \in L$, *where* L *is an extension of* K, *and let* $G \in L\{Y\} \setminus L$. *Then there exists an equivalent pc-sequence* (b_ρ) *in* K *such that* $G(b_\rho) \rightsquigarrow G(a)$.

Lemma 2.12 [ADH 2017, Lemma 6.9.1]. *Let* (a_ρ) *be a pc-sequence in* K *of d-transcendental type over* K. *Then* K *has an immediate extension* $K\langle a\rangle$ *with* a *d-transcendental over* K *and* $a_\rho \rightsquigarrow a$ *such that for any extension* L *of* K *and any* $b \in L$ *with* $a_\rho \rightsquigarrow b$, *there is a unique embedding* $K\langle a\rangle \to L$ *over* K *sending* a *to* b.

Lemma 2.13 [ADH 2017, Lemma 6.9.3]. *Let* (a_ρ) *be a pc-sequence in* K *with minimal d-polynomial* P *over* K. *Then* K *has an immediate extension* $K\langle a\rangle$ *with* $P(a) = 0$ *and* $a_\rho \rightsquigarrow a$ *such that for any extension* L *of* K *and any* $b \in L$ *with* $a_\rho \rightsquigarrow b$ *and* $P(b) = 0$, *there is a unique embedding* $K\langle a\rangle \to L$ *over* K *sending* a *to* b.

2.D. ***Constructing immediate extensions and vanishing.***

Assumption. In this subsection, the derivation induced on $\boldsymbol{k}$ is nontrivial and Γ has no least positive element.

The notion of minimal d-polynomial is not first-order, so we include here a first-order variant of Lemma 2.13 that is a special case of [ADH 2018, Lemma 5.3]. We then connect it to dominant degree in a cut. Under the assumptions above, all extensions of K are *strict* [ADH 2018, Lemma 1.3], and K is *flexible* [ADH 2018, Lemma 1.15 and Corollary 3.4]. These notions are defined in [ADH 2018] but are incidental here, and mentioned only since they occur in the corresponding lemmas of [ADH 2018].

Let $\ell \notin K$ be an element in an extension of K such that $v(\ell - K) := \{v(\ell - x) : x \in K\}$ has no largest element (equivalently, ℓ is the pseudolimit of some divergent pc-sequence in K). We say that P *vanishes at* (K, ℓ) if for all $a \in K$ and $\mathfrak{v} \in K^\times$ with $a - \ell \prec \mathfrak{v}$, $\operatorname{ddeg}_{\prec \mathfrak{v}} P_{+a} \geqslant 1$. Then $Z(K, \ell)$ denotes the set of nonzero d-polynomials over K vanishing at (K, ℓ).

Lemma 2.14 [ADH 2018, Lemma 5.3]. *Suppose that $Z(K, \ell) \neq \varnothing$ and $P \in Z(K, \ell)$ has minimal complexity. Then K has an immediate extension $K\langle f \rangle$ such that $P(f) = 0$ and $v(a - f) = v(a - \ell)$ for all $a \in K$. Moreover, if M is an extension of K and $s \in M$ satisfies $P(s) = 0$ and $v(a - s) = v(a - \ell)$ for all $a \in K$, then there is a unique embedding $K\langle f \rangle \to M$ over K sending f to s.*

Lemma 2.15 [ADH 2018, Corollary 4.6]. *Suppose that (a_ρ) is a divergent pc-sequence in K with $a_\rho \rightsquigarrow \ell$. If $P(a_\rho) \rightsquigarrow 0$, then $P \in Z(K, \ell)$.*

Lemma 2.16 [ADH 2018, Corollary 4.7]. *Suppose that (a_ρ) is a divergent pc-sequence in K with $a_\rho \rightsquigarrow \ell$. Then*

$$\operatorname{ddeg}_{\boldsymbol{a}} P = \min\{\operatorname{ddeg}_{\prec \mathfrak{v}} P_{+a} : a - \ell \prec \mathfrak{v}\}.$$

In particular, $\operatorname{ddeg}_{\boldsymbol{a}} P \geqslant 1$ *if and only if* $P \in Z(K, \ell)$.

Corollary 2.17. *Suppose that (a_ρ) is a divergent pc-sequence in K with $a_\rho \rightsquigarrow \ell$. The following conditions on P are equivalent*:

(i) *$P \in Z(K, \ell)$ and has minimal complexity in $Z(K, \ell)$*;

(ii) *P is a minimal d-polynomial of (a_ρ) over K.*

Proof. The proof is the same as that of [ADH 2017, Corollary 11.4.13], using Lemma 2.14 in place of their Lemma 11.4.8, Lemma 2.11 in place of their Lemma 11.3.8, and Lemma 2.15 in place of their Lemma 11.4.11. □

In particular, $Z(K, \ell) = \varnothing$ if and only if (a_ρ) is of d-transcendental type over K, and $Z(K, \ell) \neq \varnothing$ if and only if (a_ρ) is of d-algebraic type over K.

2.E. ***Archimedean classes and coarsening.*** For $\gamma \in \Gamma$, we let $[\gamma]$ denote its archimedean class. That is,

$$[\gamma] := \{\delta \in \Gamma : |\delta| \leqslant n|\gamma| \text{ and } |\gamma| \leqslant n|\delta| \text{ for some } n\}.$$

We order the set $[\Gamma] := \{[\gamma] : \gamma \in \Gamma\}$ by $[\delta] < [\gamma]$ if $n|\delta| < |\gamma|$ for all n. Giving the set of archimedean classes their reverse order, the map $\gamma \mapsto [\gamma]$ is a convex valuation on Γ (see [ADH 2017, §2.2] for the notion of a valuation on an abelian group and [ADH 2017, §2.4] for that of a *convex* valuation on an *ordered* abelian group). In particular, the implication $[\delta] < [\gamma] \implies [\delta + \gamma] = [\gamma]$ is often used. Then for $\phi \in K^\times$ with $\phi \not\asymp 1$, the set $\Gamma_\phi := \{\gamma : [\gamma] < [v\phi]\}$ is a convex subgroup of Γ. We will use v_ϕ, the coarsening of v by Γ_ϕ, and its corresponding dominance relation, $\preccurlyeq_\phi$, defined by

$$v_\phi : K^\times \to \Gamma/\Gamma_\phi, \quad a \mapsto va + \Gamma_\phi,$$

and $a \preccurlyeq_\phi b$ if $v_\phi(a) \geqslant v_\phi(b)$. Note that the symbols v_ϕ and $\preccurlyeq_\phi$ also appeared in [ADH 2017, §9.4], where they indicated a different coarsening of v.

We first recall how $v(P)$ changes as we additively and multiplicatively conjugate P.

Lemma 2.18 [ADH 2017, Lemma 4.5.1]. *Let $f \in K$.*

(i) *If $f \preccurlyeq 1$, then $P_{+f} \asymp P$; if $f \prec 1$, then $P_{+f} \sim P$.*

(ii) *If $f \neq 0$, then $v(P_{\times f}) \in \Gamma$ depends only on $vf \in \Gamma$.*

Item (ii) allows us to define the function $v_P : \Gamma \to \Gamma$ by $vf \mapsto v(P_{\times f})$. The main property of this function is recorded in the following lemma. Here, for $\alpha, \beta \in \Gamma$ we write $\alpha = o(\beta)$ if $[\alpha] < [\beta]$.

Lemma 2.19 [ADH 2017, Corollaries 6.1.3 and 6.1.5]. *Let $P, Q \in K\{Y\}^{\neq}$ be homogeneous of degrees m, n, respectively. For $\alpha, \beta \in \Gamma$ with $\alpha \neq \beta$, we have*

$$v_P(\alpha) - v_P(\beta) = m(\alpha - \beta) + o(\alpha - \beta).$$

It follows that for $\gamma \in \Gamma^{\neq}$,

$$v_P(\gamma) - v_Q(\gamma) = v(P) - v(Q) + (m - n)\gamma + o(\gamma),$$

and if $m > n$, then $v_P - v_Q$ is strictly increasing.

The lemmas from the rest of this subsection will play an important role in Section 6. Lemmas 2.20–2.25 are variants of lemmas from the end of [ADH 2017, §9.4]. The first two are facts about valued fields, not involving the derivation.

Lemma 2.20. *Let $f, g \in K^\times$ with $f, g \not\asymp 1$. Then $f \prec_g g$ implies $f \prec_f g$.*

Proof. From $f/g \prec_g 1$, we obtain $[vf - vg] \geqslant [vg]$. If $[vf - vg] < [vf]$, then

$$[vf] = [vf - (vf - vg)] = [vg],$$

contradicting $[vf - vg] \geqslant [vg]$. Thus $[vf - vg] \geqslant [vf]$. But since $vf - vg > 0$, we have $f/g \prec_f 1$, that is, $f \prec_f g$. □

Lemma 2.21. *Let $\phi_1, \phi_2 \in K^\times$ with $\phi_1, \phi_2 \not\asymp 1$ and $[v\phi_1] \leqslant [v\phi_2]$. Then for all $f, g \in K$, we have*

$$f \preccurlyeq_{\phi_1} g \implies f \preccurlyeq_{\phi_2} g \qquad \text{and} \qquad f \prec_{\phi_2} g \implies f \prec_{\phi_1} g.$$

In particular, $f \asymp_{\phi_1} g$ implies $f \asymp_{\phi_2} g$.

Proof. Note that for $\phi \in K^\times$ with $\phi \not\asymp 1$, $f \preccurlyeq_\phi g$ if and only if $vf - vg \in \Gamma_\phi$ or $vf > vg$, and $f \prec_\phi g$ if and only if $vf - vg > \Gamma_\phi$. Both implications then follow from $\Gamma_{\phi_1} \subseteq \Gamma_{\phi_2}$. □

Lemma 2.22. *Suppose that P is homogeneous of degree d and let $g \in K^\times$ with $g \not\asymp 1$. Then*

$$P_{\times g} \asymp_g g^d P.$$

Proof. By Lemma 2.19, $v(P_{\times g}) = vP + d\,vg + o(vg)$, so $v_g(P_{\times g}) = v_g(g^d P)$. □

Lemma 2.23. *Suppose that $g \in K^\times$ with $g \prec 1$ and $d = \operatorname{dmul} P = \operatorname{mul} P$. Then $P_{\times g} \asymp_g g^d P$.*

Proof. Since $d = \operatorname{dmul} P$, we have $P_i \preccurlyeq P_d$ for $i \geqslant d$. Since $g \prec 1$, we also have $g \prec_g 1$. Hence $g^i P_i \prec_g g^d P_d$ for $i > d$, so

$$P_{\times g} \asymp_g P_{d,\times g} \asymp_g g^d P_d$$

by Lemma 2.22. In view of $P_d \asymp P$, this yields $P_{\times g} \asymp_g g^d P$. □

Lemma 2.24. *Suppose that $g \in K^\times$ with $g \succ 1$ and $d = \operatorname{ddeg} P = \operatorname{ddeg} P_{\times g}$. Then $g P_{>d} \preccurlyeq_g P$.*

Proof. If $P_{>d} = 0$, then the result holds trivially, so assume that $P_{>d} \neq 0$. Take $i > d$ such that $P_i \asymp P_{>d}$. Then Lemma 2.22 and the fact that $g \succ 1$ give

$$(P_{\times g})_i = (P_i)_{\times g} \asymp_g g^i P_i \succcurlyeq g^{d+1} P_i \asymp g^{d+1} P_{>d},$$

so $(P_{\times g})_{>d} \succcurlyeq_g g^{d+1} P_{>d}$. Since $\operatorname{ddeg} P_{\times g} = d$, we also have $(P_{\times g})_d \succ (P_{\times g})_{>d}$. As $\operatorname{ddeg} P = d$, $P_d \asymp P$, and so

$$g^d P \asymp g^d P_d \asymp_g (P_{\times g})_d \succ (P_{\times g})_{>d} \succcurlyeq_g g^{d+1} P_{>d},$$

using Lemma 2.22 again. Hence $P \succcurlyeq_g g P_{>d}$. □

Lemma 2.25. *Let $f, g \in K^\times$ with $f, g \not\asymp 1$ and $[vf] < [vg]$. Then $P_{\times f} \asymp_g P$.*

Proof. Take d with $P_{\times f} \asymp (P_{\times f})_d$, so $P_{\times f} \asymp_f f^d P_d$ by Lemma 2.22. Then Lemma 2.21 gives $P_{\times f} \asymp_g f^d P_d$. As $[vf] < [vg]$, we get $f \asymp_g 1$, and thus $P_{\times f} \asymp_g P_d \preccurlyeq P$, so $P_{\times f} \preccurlyeq_g P$. Now, apply the same argument to $P_{\times f}$ and f^{-1} in place of P and f, using $[v(f^{-1})] = [-vf] = [vf]$, to get $P = (P_{\times f})_{\times f^{-1}} \preccurlyeq_g P_{\times f}$, and hence $P \asymp_g P_{\times f}$. □

Assumption. In the next two results, K has a monomial group $\mathfrak{M}$.

Let $\mathfrak{m}, \mathfrak{n}$ range over $\mathfrak{M}$. These two results are based on [ADH 2017, Lemma 13.2.3 and Corollary 13.2.4]:

Lemma 2.26. *Let $\mathfrak{n} \neq 1$ and $[v\mathfrak{m}] < [v\mathfrak{n}]$. Suppose that $P = Q + R$ with $R \prec_{\mathfrak{n}} P$. Then*

$$D_{P_{\times\mathfrak{m}}} = D_{Q_{\times\mathfrak{m}}}.$$

Proof. From $R \prec_{\mathfrak{n}} Q$, we get $R \prec Q$, so if $\mathfrak{m} = 1$, then $D_P = D_Q$. Now assume that $\mathfrak{m} \neq 1$. Then Lemma 2.25 gives

$$R_{\times\mathfrak{m}} \asymp_{\mathfrak{n}} R \prec_{\mathfrak{n}} Q \asymp_{\mathfrak{n}} Q_{\times\mathfrak{m}},$$

so $R_{\times\mathfrak{m}} \prec Q_{\times\mathfrak{m}}$, and hence $D_{P_{\times\mathfrak{m}}} = D_{Q_{\times\mathfrak{m}}}$. □

Corollary 2.27. *Suppose that* $\mathfrak{n} \succ 1$ *and* $\operatorname{ddeg} P = \operatorname{ddeg} P_{\times\mathfrak{n}} = d$. *Let* $Q := P_{\leqslant d}$. *Then for all* $\mathfrak{m}$ *with* $[v\mathfrak{m}] < [v\mathfrak{n}]$ *and all* $g \preccurlyeq 1$ *in* K, *we have*

$$D_{P_{+g,\times\mathfrak{m}}} = D_{Q_{+g,\times\mathfrak{m}}}.$$

Proof. Let $R := P - Q = P_{>d}$. Then Lemma 2.24 gives

$$R \preccurlyeq_{\mathfrak{n}} \mathfrak{n}^{-1} P \prec_{\mathfrak{n}} P.$$

Let $g \preccurlyeq 1$. Then $R_{+g} \asymp R$ and $P_{+g} \asymp P$ by Lemma 2.18(i). Thus we have $R_{+g} \prec_{\mathfrak{n}} P_{+g}$, so it remains to apply the previous lemma. □

3. Main results

Assuming Proposition 3.1, we prove here the main results of this paper concerning the uniqueness of maximal immediate extensions, the relationship between d-algebraic maximality and d-henselianity, and the existence and uniqueness of d-henselizations. The proof of Proposition 3.1 is given in Section 7.

Proposition 3.1. *Suppose that* K *is asymptotic*, Γ *is divisible*, *and* $\boldsymbol{k}$ *is* r*-linearly surjective. Let* (a_ρ) *be a pc-sequence in* K *with minimal d-polynomial* G *over* K *of order at most* r. *Then* $\operatorname{ddeg}_{\boldsymbol{a}} G = 1$.

3.A. ***Removing divisibility.*** In the next lemmas, we construe the algebraic closure K^{ac} of K as a valued differential field extension of K: the derivation of K extends uniquely to K^{ac} [ADH 2017, Lemma 1.9.2] and we equip K^{ac} with any valuation extending that of K. This determines K^{ac} as a valued differential field extension of K up to isomorphism over K, with value group the divisible hull $\mathbb{Q}\Gamma$ of Γ and residue field the algebraic closure $\boldsymbol{k}^{\mathrm{ac}}$ of $\boldsymbol{k}$. If K is henselian, then its valuation extends uniquely to K^{ac}; see [ADH 2017, Proposition 3.3.11]. By [ADH 2017, Proposition 6.2.1], K^{ac} has small derivation.

Lemma 3.2. *Suppose that* K *is henselian and the derivation on* $\boldsymbol{k}$ *is nontrivial. Let* (a_ρ) *be a pc-sequence in* K *with minimal d-polynomial* P *over* K. *Then* P *remains a minimal d-polynomial of* (a_ρ) *over the algebraic closure* K^{ac} *of* K.

Proof. We may suppose that (a_ρ) is divergent in K, the other case being trivial. Then (a_ρ) is still divergent in K^{ac}: If it had a pseudolimit $a \in K^{\mathrm{ac}}$, then we would have $Q(a_\rho) \rightsquigarrow 0$, where $Q \in K[Y]$ is the minimum polynomial of a over K (see [ADH

2017, Proposition 3.2.1]). But since K is henselian, it is algebraically maximal (see [ADH 2017, Corollary 3.3.21]), and then (a_ρ) would have a pseudolimit in K.

Now suppose to the contrary that Q is a minimal d-polynomial of (a_ρ) over K^{ac} with $c(Q) < c(P)$. Take an extension $L \subseteq K^{\mathrm{ac}}$ of K with $[L : K] = n$ and $Q \in L\{Y\}$. Then as K is henselian,

$$[L : K] = [\Gamma_L : \Gamma] \cdot [\boldsymbol{k}_L : \boldsymbol{k}]$$

(see [ADH 2017, Corollary 3.3.49]), so we have a valuation basis $\mathcal{B} = \{e_1, \dots, e_n\}$ of L over K (see [ADH 2017, Proposition 3.1.7]). That is, $\mathcal{B}$ is a basis of L over K, and for all $a_1, \dots, a_n \in K$,

$$v\left(\sum_{i=1}^{n} a_i e_i\right) = \min_{1 \leqslant i \leqslant n} v(a_i e_i).$$

Then by expressing the coefficients of Q in terms of the valuation basis,

$$Q(Y) = \sum_{i=1}^{n} R_i(Y) \cdot e_i,$$

where $R_i \in K\{Y\}$ for $1 \leqslant i \leqslant n$.

Since Q is a minimal d-polynomial of (a_ρ) over K^{ac}, by Lemma 2.13 we have an immediate extension $K^{\mathrm{ac}}\langle a\rangle$ of K^{ac} with $a_\rho \rightsquigarrow a$ and $Q(a) = 0$. Then by Lemma 2.11, we have a pc-sequence (b_ρ) in K equivalent to (a_ρ) such that $Q(b_\rho) \rightsquigarrow Q(a) = 0$. Finally, after passing to a cofinal subsequence, we can assume that we have i with $Q(b_\rho) \asymp R_i(b_\rho) \cdot e_i$ for all ρ. Then $R_i(b_\rho) \rightsquigarrow 0$ and $c(R_i) < c(P)$, contradicting the minimality of P. □

Since minimal d-polynomials are irreducible, note that a corollary of this lemma is that minimal d-polynomials over henselian K (with nontrivial induced derivation on $\boldsymbol{k}$) are absolutely irreducible. We can now replace the divisibility assumption in the main proposition with that of henselianity.

Proposition 3.3. *Suppose that K is asymptotic and henselian, and that $\boldsymbol{k}$ is linearly surjective. Let (a_ρ) be a pc-sequence in K with minimal d-polynomial G over K. Then* $\operatorname{ddeg}_{\boldsymbol{a}} G = 1$.

Proof. By the previous lemma, G remains a minimal d-polynomial of (a_ρ) over K^{ac}. Note that the value group of K^{ac} is divisible, and its differential residue field is linearly surjective, as an algebraic extension of $\boldsymbol{k}$ [ADH 2017, Corollary 5.4.3]. But then $\operatorname{ddeg}_{\boldsymbol{a}_{K^{\mathrm{ac}}}} G = 1$ by Proposition 3.1, and so $\operatorname{ddeg}_{\boldsymbol{a}} G = 1$ by Lemma 2.10(vii). □

3.B. *Main results.* Recall from [van den Dries and Pynn-Coates 2019] that K has the *differential-henselian configuration property* (*dh-configuration property* for short) if for every divergent pc-sequence (a_ρ) in K with minimal d-polynomial G over K, we have $\operatorname{ddeg}_{\boldsymbol{a}} G = 1$. In that paper, we showed that several results follow from the dh-configuration property. In particular, Lemma 3.4, Theorem 3.6, and Corollary 3.8 follow immediately in view of Proposition 3.3. For readability, we include their proofs from that paper verbatim. Theorems 3.5 and 3.7 require minor modifications involving henselizations.

The next lemma has the same proof as [van den Dries and Pynn-Coates 2019, Lemma 4.1], except for using Proposition 3.3.

Lemma 3.4. *Suppose K is asymptotic and henselian, and $\boldsymbol{k}$ is linearly surjective. Let (a_ρ) be a pc-sequence in K with minimal d-polynomial G over K. Let L be a d-algebraically maximal extension of K such that $\boldsymbol{k}_L$ is linearly surjective. Then there is $b \in L$ with $a_\rho \rightsquigarrow b$ and $G(b) = 0$.*

Proof. Note that L is d-henselian by [ADH 2017, Theorem 7.0.1]. Since L is d-algebraically maximal and the derivation of $\boldsymbol{k}_L$ is nontrivial, every pc-sequence in L of d-algebraic type over L has a pseudolimit in L by Lemma 2.13. Thus we get $a \in L$ with $a_\rho \rightsquigarrow a$. Passing to an equivalent pc-sequence we arrange that $G(a_\rho) \rightsquigarrow 0$. With $\gamma_\rho = v(a_{\rho+1} - a_\rho) = v(a - a_\rho)$, eventually, Proposition 3.3 gives $g_\rho \in K$ with $v(g_\rho) = \gamma_\rho$ and $\operatorname{ddeg} G_{+a_\rho, \times g_\rho} = 1$, eventually. By Corollary 2.3, $\operatorname{ddeg} G_{+a, \times g_\rho} = 1$, eventually. We have $G(a + Y) = G(a) + A(Y) + R(Y)$, where A is linear and homogeneous and all monomials in R have degree $\geqslant 2$, and so

$$G_{+a, \times g_\rho}(Y) = G(a) + A_{\times g_\rho}(Y) + R_{\times g_\rho}(Y).$$

Now $\operatorname{ddeg} G_{+a, \times g_\rho} = 1$ eventually, so $v(G(a)) \geqslant v_A(\gamma_\rho) < v_R(\gamma_\rho)$ eventually. With $v(G, a)$ as in [ADH 2017, §7.3] we get $v(G, a) > v(a - a_\rho)$, eventually. Then [ADH 2017, Lemma 7.3.1] gives $b \in L$ with $v_L(a - b) = v(G, a)$ and $G(b) = 0$, so $v_L(a - b) > v(a - a_\rho)$, eventually. Thus $a_\rho \rightsquigarrow b$. □

For the next result, we copy the proof of [van den Dries and Pynn-Coates 2019, Theorem 4.2], except for an argument involving henselizations.

Theorem 3.5. *Suppose that K is asymptotic and $\boldsymbol{k}$ is linearly surjective. Then any two maximal immediate extensions of K are isomorphic over K. Also, any two d-algebraically maximal d-algebraic immediate extensions of K are isomorphic over K.*

Proof. Let L_0 and L_1 be maximal immediate extensions of K. By Zorn's lemma we have a maximal isomorphism $\mu : F_0 \to F_1$ over K between valued differential subfields $F_i \supseteq K$ of L_i for $i = 0, 1$, where "maximal" means that μ does not extend to an isomorphism between strictly larger such valued differential subfields. First, F_i is asymptotic by [ADH 2017, Lemmas 9.4.2 and 9.4.5], and $\boldsymbol{k}_{F_i}$ is linearly surjective, as F_i is an immediate extension of K for $i = 0, 1$. Next, F_i must be henselian, since its henselization in L_i is algebraic over F_i, and thus a valued differential subfield of L_i for $i = 0, 1$. Now suppose towards a contradiction that

$F_0 \neq L_0$ (equivalently, $F_1 \neq L_1$). Then F_0 is not spherically complete, so we have a divergent pc-sequence (a_ρ) in F_0.

Suppose that (a_ρ) is of d-transcendental type over F_0. The spherical completeness of L_0 and L_1 then yields $f_0 \in L_0$ and $f_1 \in L_1$ such that $a_\rho \rightsquigarrow f_0$ and $\mu(a_\rho) \rightsquigarrow f_1$. Hence by Lemma 2.12 we obtain an isomorphism $F_0\langle f_0\rangle \to F_1\langle f_1\rangle$ extending μ, contradicting the maximality of μ.

Suppose that (a_ρ) is of d-algebraic type over F_0, with minimal d-polynomial G over F_0. Then Lemma 3.4 gives $f_0 \in L_0$ with $a_\rho \rightsquigarrow f_0$ and $G(f_0) = 0$, and $f_1 \in L_1$ with $\mu(a_\rho) \rightsquigarrow f_1$ and $\mu(G)(f_1) = 0$, where we extend μ to the differential ring isomorphism $\mu : F_0\{Y\} \to F_1\{Y\}$ with $Y \mapsto Y$. Now Lemma 2.13 gives an isomorphism $F_0\langle f_0\rangle \to F_1\langle f_1\rangle$ extending μ, and we have again a contradiction. Thus $F_0 = L_0$ and hence $F_1 = L_1$ as well.

The proof of the second statement is the same, using only Lemma 2.13. □

By Proposition 3.3 and [van den Dries and Pynn-Coates 2019, Theorem 4.3], we can now remove monotonicity from [ADH 2017, Theorem 7.0.3] without any further assumptions.

Theorem 3.6. *If K is asymptotic and d-henselian, then it is d-algebraically maximal.*

Proof. Let (a_ρ) be a pc-sequence in K with minimal d-polynomial G over K. Towards a contradiction, assume that (a_ρ) is divergent in K. Then Lemma 2.13 shows that G has order at least 1 (since K is henselian) and provides a proper immediate extension $K\langle a\rangle$ of K with $a_\rho \rightsquigarrow a$. Replacing (a_ρ) by an equivalent pc-sequence in K, we arrange that $G(a_\rho) \rightsquigarrow 0$.

By Proposition 3.3, $\operatorname{ddeg}_{\boldsymbol{a}} G = 1$. Taking $g_\rho \in K$ with $v(g_\rho) = \gamma_\rho$ we have $\operatorname{ddeg} G_{+a_\rho, \times g_\rho} = 1$, eventually. By removing some initial terms of the sequence, we arrange that this holds for all ρ and that $v(a - a_\rho) = \gamma_\rho$ for all ρ. By d-henselianity, we have $z_\rho \in K$ with $G(z_\rho) = 0$ and $a_\rho - z_\rho \preccurlyeq g_\rho$. From $a - a_\rho \asymp g_\rho$ we get $a - z_\rho \preccurlyeq g_\rho$.

Let $r \geqslant 1$ be the order of G. By [ADH 2017, Lemma 2.2.19], (γ_ρ) is cofinal in $v(a - K)$, so there are indices $\rho_0 < \rho_1 < \cdots < \rho_{r+1}$ such that $a - z_{\rho_j} \prec a - z_{\rho_i}$, for $0 \leqslant i < j \leqslant r + 1$. Then

$$z_{\rho_i} - z_{\rho_{i-1}} \asymp a - z_{\rho_{i-1}} \succ a - z_{\rho_i} \asymp z_{\rho_{i+1}} - z_{\rho_i} \quad \text{for } 1 \leqslant i \leqslant r.$$

We have $a - z_{\rho_{r+1}} \prec a - z_{\rho_0} \preccurlyeq g_{\rho_0}$, so $z_{\rho_0} - z_{\rho_{r+1}} \preccurlyeq g_{\rho_0}$, and thus also $a_{\rho_0} - z_{\rho_{r+1}} \preccurlyeq g_{\rho_0}$. Hence

$$\operatorname{ddeg} G_{+z_{\rho_{r+1}}, \times g_{\rho_0}} = \operatorname{ddeg} G_{+z_{\rho_0}, \times g_{\rho_0}} = \operatorname{ddeg} G_{+a_{\rho_0}, \times g_{\rho_0}} = 1$$

by Corollary 2.3. Thus, with z_{ρ_i} in the role of y_i, for $0 \leqslant i \leqslant r + 1$, and g_{ρ_0} in the role of g, we have reached a contradiction with [ADH 2017, Lemma 7.5.5]. □

The following generalizes [van den Dries and Pynn-Coates 2019, Theorem 1.3], removing the assumption on the value group. Its proof is the same, except for the use of the henselization.

Theorem 3.7. *Suppose that K is asymptotic and $\boldsymbol{k}$ is linearly surjective. Then K has a d-henselization which has no proper differential subfield containing K that is d-henselian. In particular*, *any two d-henselizations of K are isomorphic over K.*

Proof. We can assume that K is henselian, as K has a henselization that embeds (uniquely) over K into any d-henselian extension of K. By [ADH 2017, Corollary 9.4.11] we have an immediate asymptotic d-henselian extension K^{dh} of K that is d-algebraic over K and has no proper differential subfield containing K that is d-henselian.

Let L be an immediate d-henselian extension of K; then L is asymptotic by [ADH 2017, Lemmas 9.4.2 and 9.4.5]. To see that K^{dh} embeds into L over K, use an argument similar to that in the proof of Theorem 3.5, as K^{dh} and L are d-algebraically maximal by Theorem 3.6. Thus K^{dh} is a d-henselization of K and any d-henselization of K is isomorphic over K to K^{dh}. □

In fact, the argument shows that K^{dh} as in the proof of Theorem 3.7 embeds over K into any (not necessarily immediate) asymptotic d-henselian extension of K. This corollary has the same proof as [van den Dries and Pynn-Coates 2019, Corollary 4.6].

Corollary 3.8. *If K is asymptotic and $\boldsymbol{k}$ is linearly surjective*, *then any immediate d-henselian extension of K that is d-algebraic over K is a d-henselization of K.*

Proof. Let K^{dh} be the d-henselization of K from the proof of Theorem 3.7. Then K^{dh} is asymptotic, so is d-algebraically maximal by Theorem 3.6. Hence any embedding $K^{\mathrm{dh}} \to L$ over K into an immediate d-algebraic extension L of K is surjective. □

3.C. ***Additional results.*** We also record versions of the above results relativized to d-polynomials of a given order. In these results, we assume that Γ is divisible but not that K is henselian. The proofs are the same as above, except for using Proposition 3.1 in place of Proposition 3.3 and not needing to use henselizations.

To state these results, we make some definitions. If E and F are differential fields, then we say that F is *r-differentially algebraic* (*r-d-algebraic* for short) over E if for each $a \in F$ there are $a_1, \ldots, a_n \in F$ such that $a \in E\langle a_1, \ldots, a_n\rangle$ and, for $i = 0, \ldots, n-1$, a_{i+1} is d-algebraic over $E\langle a_1, \ldots, a_i\rangle$ with minimal annihilator of order at most r. It is routine to prove that if L is r-d-algebraic over F and F is r-d-algebraic over E, then L is r-d-algebraic over E.

Definition. We call K *r-differential-algebraically maximal* (*r-d-algebraically maximal* for short) if it has no proper immediate r-d-algebraic extension.

By Zorn, K has an immediate r-d-algebraic extension that is r-d-algebraically maximal. Note that K is d-algebraically maximal if and only if it is r-d-algebraically maximal for all r. In addition, if the derivation induced on $\boldsymbol{k}$ is nontrivial, then by Lemmas 2.11 and 2.13 K is r-d-algebraically maximal if and only if every pc-sequence in K with minimal d-polynomial over K of order at most r has a pseudolimit in K.

Note that K being 0-d-algebraically maximal means that it has no proper immediate valued differential field extension with small derivation that is algebraic over K. Since each algebraic field extension of K, given any valuation extending that of K and the unique derivation extending that of K, has small derivation [ADH 2017, Proposition 6.2.1], K is 0-d-algebraically maximal if and only if it is algebraically maximal *as a valued field*. Thus the results below for $r = 0$ follow from the corresponding results for valued fields and hence we may assume that $r \geqslant 1$, so the derivation induced on $\boldsymbol{k}$ is nontrivial when $\boldsymbol{k}$ is r-linearly surjective, as was used in the preceding proofs.

Theorem 3.9. *If K is asymptotic, Γ is divisible, and $\boldsymbol{k}$ is r-linearly surjective, then any two r-d-algebraically maximal r-d-algebraic immediate extensions of K are isomorphic over K.*

The proof of [ADH 2017, Theorem 7.0.1] shows that if K is r-d-algebraically maximal and $\boldsymbol{k}$ is r-linearly surjective, then K is r-d-henselian. Conversely:

Theorem 3.10. *If K is asymptotic and r-d-henselian, and Γ is divisible, then K is r-d-algebraically maximal.*

We say an extension L of K is an *r-differential-henselization* (*r-d-henselization* for short) of K if it is an immediate asymptotic r-d-henselian extension of K that embeds over K into any asymptotic r-d-henselian extension of K. In the next proof, instead of using [ADH 2017, Corollary 9.4.11] we let K^{dh} be any r-d-algebraically maximal immediate r-d-algebraic extension of K, since K^{dh} is r-d-henselian and no proper differential subfield of K^{dh} containing K is r-d-henselian by Theorem 3.10.

Theorem 3.11. *If K is asymptotic, Γ is divisible, and $\boldsymbol{k}$ is r-linearly surjective, then K has an r-d-henselization, and any two r-d-henselizations of K are isomorphic over K.*

Corollary 3.12. *If K is asymptotic, Γ is divisible, and $\boldsymbol{k}$ is r-linearly surjective, then any immediate r-d-henselian extension of K that is r-d-algebraic over K is an r-d-henselization of K.*

4. Newton diagrams

We develop a differential newton diagram method for valued differential fields with small derivation. This approach is closely modelled on the differential newton diagram method for a certain class of asymptotic fields developed in [ADH 2017, §13.5]. In Section 4.A, we connect this to dominant degree in a cut, adapting two lemmas from [ADH 2017, §13.6]. The assumption of divisible value group allows us to use the equalizer theorem, which underlies this method:

Theorem 4.1 [ADH 2017, Theorem 6.0.1]. *Let $P, Q \in K\{Y\}^{\neq}$ be homogeneous of degrees m, n, respectively, with $m > n$. If $(m-n)\Gamma = \Gamma$, then there exists a unique $\alpha \in \Gamma$ such that $v_P(\alpha) = v_Q(\alpha)$.*

Assumption. In this section, K has a monomial group $\mathfrak{M}$.

Let P range over $K\{Y\}^{\neq}$, f and g over K, and $\mathfrak{m}$ and $\mathfrak{n}$ over $\mathfrak{M}$. For $f \in K^{\times}$, let $\mathfrak{d}_f$ be the unique monomial with $\mathfrak{d}_f \asymp f$ and $u_f := f/\mathfrak{d}_f \asymp 1$. The next lemma plays a subtle but important role in this section and the next. It is based on [ADH 2017, Lemma 13.5.4], but with a somewhat different proof due to the differences between dominant parts and newton polynomials (the appropriate analogue of dominant parts in that setting).

Lemma 4.2. *Suppose that Γ has no least positive element and $f \preccurlyeq \mathfrak{m}$. If $f \prec \mathfrak{m}$, let $u := 0$; if $f \asymp \mathfrak{m}$, let $u := u_f$. Then*

$$\operatorname{ddeg}_{\prec\mathfrak{m}} P_{+f} = \operatorname{mul}(D_{P_{\times\mathfrak{m}}})_{+\bar{u}}.$$

In particular, $\operatorname{ddeg}_{\prec\mathfrak{m}} P = \operatorname{dmul} P_{\times\mathfrak{m}}$.

Proof. For $\mathfrak{n} \prec \mathfrak{m}$, let $\mathfrak{e} = \mathfrak{n}\mathfrak{m}^{-1} \in \mathfrak{M}$. Then

$$P_{+f,\times\mathfrak{n}} = P_{\times\mathfrak{m},+\mathfrak{m}^{-1}f,\times\mathfrak{e}},$$

so by replacing P with $P_{\times\mathfrak{m}}$ and f with $\mathfrak{m}^{-1}f$, we may assume that $\mathfrak{m} = 1$. Set $Q := P_{+f}$, so by Lemma 2.2(i), $D_Q = (D_P)_{+\bar{f}} = (D_P)_{+\bar{u}}$. Thus $\operatorname{mul}(D_P)_{+\bar{u}} = \operatorname{dmul} Q$, so it remains to show

$$\operatorname{ddeg}_{\prec 1} Q = \operatorname{dmul} Q.$$

First, we have $\operatorname{ddeg}_{\prec 1} Q \leqslant \operatorname{dmul} Q$ by Corollary 2.4. For the other direction, let $d := \operatorname{dmul} Q$. We have $v(Q_d) < v(Q_i)$ for all $i < d$, so take $g \prec 1$ with vg small enough that

$$v(Q_d) + (d+1)\, vg < v(Q_i) \quad \text{for all } i < d.$$

It follows that

$$v(Q_d) + d\, vg + o(vg) < v(Q_i) + i\, vg + o(vg) \quad \text{for all } i < d,$$

so $v(Q_{d,\times g}) < v(Q_{i,\times g})$ for all $i < d$ by Lemma 2.19. Hence $\operatorname{dmul} Q_{\times g} \geqslant d$. But

$$\operatorname{ddeg}_{\prec 1} Q = \max\{\operatorname{dmul} Q_{\times g} : g \prec 1\}$$

by Lemma 2.5, so $\operatorname{ddeg}_{\prec 1} Q \geqslant d$, as desired. □

We call $y \in K^\times$ an *approximate zero of* P if, for $\mathfrak{m} := \mathfrak{d}_y$ and $u := u_y$, $D_{P_{\times\mathfrak{m}}}(\bar{u}) = 0$. If y is an approximate zero of P, we define its *multiplicity* to be $\operatorname{mul}(D_{P_{\times\mathfrak{m}}})_{+\bar{u}}$. We call $\mathfrak{m}$ an *algebraic starting monomial for* P if $D_{P_{\times\mathfrak{m}}}$ is not homogeneous. In particular, if $\mathfrak{m}$ is an algebraic starting monomial for P, then $\operatorname{ddeg} P_{\times\mathfrak{m}} \geqslant 1$. Note that $\mathfrak{m}$ is an algebraic starting monomial for P if and only if $\mathfrak{m}/\mathfrak{n}$ is an algebraic starting monomial for $P_{\times\mathfrak{n}}$. By Corollary 2.4, P has at most $\deg P - \operatorname{mul} P$ algebraic starting monomials.

Assumption. In the rest of this section, Γ is divisible.

The existence of algebraic starting monomials is an easy corollary of the equalizer theorem, and is crucial to what follows. It corresponds to [ADH 2017, Corollary 13.5.6].

Lemma 4.3. *Let $P, Q \in K\{Y\}^{\neq}$ be homogeneous of different degrees. Then there exists a unique $\mathfrak{m}$ such that $D_{(P+Q)_{\times\mathfrak{m}}}$ is not homogeneous.*

Proof. By Theorem 4.1, there is a unique $\mathfrak{m}$ such that $P_{\times\mathfrak{m}} \asymp Q_{\times\mathfrak{m}}$. Then

$$D_{(P+Q)_{\times\mathfrak{m}}} = D_{P_{\times\mathfrak{m}}+Q_{\times\mathfrak{m}}} = D_{P_{\times\mathfrak{m}}} + D_{Q_{\times\mathfrak{m}}}$$

by Lemma 2.1(ii), so $D_{(P+Q)_{\times\mathfrak{m}}}$ is not homogeneous. For $\mathfrak{n} \neq \mathfrak{m}$, we have $D_{(P+Q)_{\times\mathfrak{n}}} = D_{P_{\times\mathfrak{n}}}$ or $D_{(P+Q)_{\times\mathfrak{n}}} = D_{Q_{\times\mathfrak{n}}}$ by Lemma 2.1(i), since $P_{\times\mathfrak{n}} \succ Q_{\times\mathfrak{n}}$ or $P_{\times\mathfrak{n}} \prec Q_{\times\mathfrak{n}}$. □

For P and Q as in Lemma 4.3, we let $\mathfrak{e}(P, Q)$ denote the unique monomial which that lemma yields and call it the *equalizer for P, Q*. We are interested in the case that these two d-polynomials are homogeneous parts of the same d-polynomial. Let $J := \{j \in \mathbb{N} : P_j \neq 0\}$ and note that $\operatorname{ddeg} P_{\times\mathfrak{m}} \in J$ for all $\mathfrak{m}$. For distinct $i, j \in J$, let $\mathfrak{e}(P, i, j) := \mathfrak{e}(P_i, P_j)$, and so any algebraic starting monomial for P is of the form $\mathfrak{e}(P, i, j)$ for some distinct $i, j \in J$.

In the next two results, let $\mathcal{E} \subseteq K^\times$ be $\preccurlyeq$-closed. Recall this means that $\mathcal{E} \neq \varnothing$ and $f \in \mathcal{E}$ whenever $0 \neq f \preccurlyeq g$ with $g \in \mathcal{E}$. This result corresponds to [ADH 2017, Proposition 13.5.7]:

Proposition 4.4. *There exist $i_0, \ldots, i_n \in J$ and equalizers*

$$\mathfrak{e}(P, i_0, i_1) \prec \mathfrak{e}(P, i_1, i_2) \prec \cdots \prec \mathfrak{e}(P, i_{n-1}, i_n)$$

with $\operatorname{mul} P = i_0 < \cdots < i_n = \operatorname{ddeg}_{\mathcal{E}} P$ *such that*

(i) *the algebraic starting monomials for P in $\mathcal{E}$ are the $\mathfrak{e}(P, i_m, i_{m+1})$ for $m < n$;*

(ii) *for $m < n$ and $\mathfrak{m} = \mathfrak{e}(P, i_m, i_{m+1})$, we have*

$$\operatorname{dmul} P_{\times\mathfrak{m}} = i_m \quad \textit{and} \quad \operatorname{ddeg} P_{\times\mathfrak{m}} = i_{m+1}.$$

Proof. Let i, j range over J and $d := \operatorname{ddeg}_{\mathcal{E}} P$. Then $\operatorname{mul} P \leqslant d \leqslant \deg P$, and we proceed by induction on $d - \operatorname{mul} P$. If $d = \operatorname{mul} P$, then for all $\mathfrak{m} \in \mathcal{E}$, $D_{P_{\times\mathfrak{m}}}$ is homogeneous of degree d, so there is no algebraic starting monomial for P in $\mathcal{E}$.

Now assume that $d > \operatorname{mul} P$ and take $i < d$ such that $\mathfrak{e} := \mathfrak{e}(P, i, d) \succcurlyeq \mathfrak{e}(P, j, d)$ for all $j < d$. First, we show that $\mathfrak{e} \in \mathcal{E}$. We have $P_{d,\times\mathfrak{e}} \asymp P_{i,\times\mathfrak{e}}$ by the previous lemma, so $v_{P_d}(v\mathfrak{e}) = v_{P_i}(v\mathfrak{e})$. By Lemma 2.19, the function $v_{P_d} - v_{P_i}$ is strictly increasing, so for any $g \prec \mathfrak{e}$, we have $v_{P_d}(vg) > v_{P_i}(vg)$, that is, $P_{d,\times g} \prec P_{i,\times g}$. Hence $\operatorname{ddeg} P_{\times g} < d$. To obtain $\mathfrak{e} \in \mathcal{E}$, take $g \in \mathcal{E}$ with $\operatorname{ddeg} P_{\times g} = d$, so $\mathfrak{e} \preccurlyeq g$.

Next, we show that $\operatorname{ddeg} P_{\times\mathfrak{e}} = d$. If $\operatorname{ddeg} P_{\times\mathfrak{e}} = j < d$, then $P_{d,\times\mathfrak{e}} \prec P_{j,\times\mathfrak{e}}$. By Lemma 2.19 again, the function $v_{P_d} - v_{P_j}$ is strictly increasing, so it follows that $\mathfrak{e} \prec \mathfrak{e}(P, j, d)$, contradicting the maximality of $\mathfrak{e}$.

From this and $P_{i,\times\mathfrak{e}} \asymp P_{d,\times\mathfrak{e}}$, we get $(D_{P_{\times\mathfrak{e}}})_i = D_{P_{i,\times\mathfrak{e}}} \neq 0$ and $(D_{P_{\times\mathfrak{e}}})_d = D_{P_{d,\times\mathfrak{e}}} \neq 0$, and hence $\mathfrak{e}$ is an algebraic starting monomial for P. In fact, $\mathfrak{e}$ is the largest algebraic starting monomial for P in $\mathcal{E}$. Suppose to the contrary that $\mathfrak{n} \in \mathcal{E}$ is an algebraic starting monomial for P with $\mathfrak{n} \succ \mathfrak{e}$. Then $d = \operatorname{ddeg} P_{\times\mathfrak{e}} \leqslant \operatorname{ddeg} P_{\times\mathfrak{n}}$ by Corollary 2.4, so $\operatorname{ddeg} P_{\times\mathfrak{n}} = d$. It follows that $\mathfrak{n} = \mathfrak{e}(P, j, d)$ for some $j < d$, contradicting the maximality of $\mathfrak{e}$.

If $i > \operatorname{dmul} P_{\times\mathfrak{e}}$, then for $j := \operatorname{dmul} P_{\times\mathfrak{e}}$, the uniqueness in Lemma 4.3 yields $\mathfrak{e}(P, j, d) = \mathfrak{e}$. By replacing i with j, we assume that $i = \operatorname{dmul} P_{\times\mathfrak{e}}$. Then by Lemma 4.2, we also have $\operatorname{ddeg}_{\prec\mathfrak{e}} P = i$. To complete the proof, we apply the inductive assumption with $\{g \in K^{\times} : g \prec \mathfrak{e}\}$ replacing $\mathcal{E}$. □

The tuple $(i_0, \dots, i_n)$ from Proposition 4.4 is uniquely determined by K, P, and $\mathcal{E}$. Note that if $\operatorname{mul} P = \operatorname{ddeg}_{\mathcal{E}} P$, then $n = 0$ and the tuple is $(\operatorname{mul} P)$. For $1 \leqslant m \leqslant n$, set $\mathfrak{e}_m := \mathfrak{e}(P, i_{m-1}, i_m)$. We now show how $\operatorname{dmul} P_{\times g}$ and $\operatorname{ddeg} P_{\times g}$ behave as g ranges over $\mathcal{E}$. This follows from Proposition 4.4 in exactly the same way as [ADH 2017, Corollary 13.5.8] follows from [ADH 2017, Proposition 13.5.7], using Corollary 2.4 instead of [ADH 2017, Corollary 11.2.5].

Corollary 4.5. *Suppose that* $\operatorname{mul} P \neq \operatorname{ddeg}_{\mathcal{E}} P$, *so* $n \geqslant 1$. *Let g range over $\mathcal{E}$. Then* $\operatorname{dmul} P_{\times g}$ *and* $\operatorname{ddeg} P_{\times g}$ *are in* $\{i_0, \dots, i_n\}$ *and we have*

$$\begin{aligned}
\operatorname{dmul} P_{\times g} = i_0 &\iff g \preccurlyeq \mathfrak{e}_1;\\
\operatorname{ddeg} P_{\times g} = i_0 &\iff g \prec \mathfrak{e}_1;\\
\operatorname{dmul} P_{\times g} = i_m &\iff \mathfrak{e}_m \prec g \preccurlyeq \mathfrak{e}_{m+1} \qquad (1 \leqslant m < n);\\
\operatorname{ddeg} P_{\times g} = i_m &\iff \mathfrak{e}_m \preccurlyeq g \prec \mathfrak{e}_{m+1} \qquad (1 \leqslant m < n);\\
\operatorname{dmul} P_{\times g} = i_n &\iff \mathfrak{e}_n \prec g;\\
\operatorname{ddeg} P_{\times g} = i_n &\iff \mathfrak{e}_n \preccurlyeq g.
\end{aligned}$$

4.A. ***Application to dominant degree in a cut.*** The first lemma is easily adapted from [ADH 2017, Lemma 13.6.17] and its corollary corresponds to [ADH 2017, Corollary 13.6.18].

Lemma 4.6. *Suppose that* (a_ρ) *is a pc-sequence in* K *with* $a_\rho \rightsquigarrow 0$. *Let*

$$\mathcal{E}_{\boldsymbol{a}} := \{g \in K^\times : g \prec a_\rho, \text{ eventually}\}.$$

(i) *If* $\mathcal{E}_{\boldsymbol{a}} \neq \varnothing$, *then* $\operatorname{ddeg}_{\boldsymbol{a}} P = \operatorname{ddeg}_{\mathcal{E}_{\boldsymbol{a}}} P$.

(ii) *If* $\mathcal{E}_{\boldsymbol{a}} = \varnothing$, *then* $\operatorname{ddeg}_{\boldsymbol{a}} P = \operatorname{mul} P$.

Proof. Set $\gamma_\rho := v(a_{\rho+1} - a_\rho)$. By removing some initial terms, we may assume that γ_ρ is strictly increasing and $v(a_\rho) = \gamma_\rho \in \Gamma$ for all ρ. Then by Lemma 2.6,

$$\operatorname{ddeg}_{\geqslant \gamma_\rho} P_{+a_\rho} = \operatorname{ddeg}_{\geqslant \gamma_\rho} P = \operatorname{ddeg} P_{\times a_\rho},$$

so $\operatorname{ddeg}_{\boldsymbol{a}} P$ is the eventual value of $\operatorname{ddeg} P_{\times a_\rho}$. If P is homogeneous, then $\operatorname{ddeg} P_{\times g} = \deg P = \operatorname{mul} P$ for all $g \in K^\times$, so the statements about $\operatorname{ddeg}_{\boldsymbol{a}} P$ are immediate.

Suppose now that P is not homogeneous, so $\operatorname{mul} P < \deg P$. If $\mathcal{E}_{\boldsymbol{a}} \neq \varnothing$, we use Corollary 4.5 with $K^\times$ in the role of $\mathcal{E}$, so we have the tuple $(i_0, \dots, i_n)$ with $i_n = \deg P$. By removing further initial terms, we may assume that $\operatorname{ddeg} P_{\times a_\rho}$ is constant. If $\operatorname{ddeg} P_{\times a_\rho} = i_0$, then $a_\rho \prec \mathfrak{e}_1$. Thus for any $g \in \mathcal{E}_{\boldsymbol{a}}$, we have $g \prec \mathfrak{e}_1$, and hence $\operatorname{ddeg}_{\mathcal{E}_{\boldsymbol{a}}} P = i_0$. If $\operatorname{ddeg} P_{\times a_\rho} = i_m$ for any $1 \leqslant m \leqslant n$, then $\mathfrak{e}_m \preccurlyeq a_\rho$. As γ_ρ is strictly increasing, $\mathfrak{e}_m \prec a_\rho$ for all ρ, so $\mathfrak{e}_m \in \mathcal{E}_{\boldsymbol{a}}$. Hence $\operatorname{ddeg}_{\mathcal{E}_{\boldsymbol{a}}} P \geqslant \operatorname{ddeg} P_{\times \mathfrak{e}_m} = i_m$. But by Corollary 2.4, $\operatorname{ddeg} P_{\times a_\rho} \geqslant \operatorname{ddeg}_{\mathcal{E}_{\boldsymbol{a}}} P$, so $\operatorname{ddeg}_{\mathcal{E}_{\boldsymbol{a}}} P = i_m$.

If $\mathcal{E}_{\boldsymbol{a}} = \varnothing$, let $i_0 := \operatorname{mul} P$. Then for all $i > i_0$, by Lemma 2.19,

$$v_{P_i}(\gamma_\rho) - v_{P_{i_0}}(\gamma_\rho) = v(P_i) - v(P_{i_0}) + (i - i_0)\gamma_\rho + o(\gamma_\rho).$$

As γ_ρ is cofinal in Γ, we thus have $v_P(\gamma_\rho) = v_{P_{i_0}}(\gamma_\rho) < v_{P_i}(\gamma_\rho)$, eventually, for all $i > i_0$, and so $\operatorname{ddeg} P_{\times a_\rho} = i_0$, eventually. □

With (a_ρ) and $\mathcal{E}_{\boldsymbol{a}}$ as in the above lemma, if $\mathcal{E}_{\boldsymbol{a}} = \varnothing$, then (a_ρ) is in fact a cauchy sequence in K (see [ADH 2017, §2.2]), since γ_ρ is cofinal in Γ; this is not used later.

Corollary 4.7. *Suppose* (b_ρ) *is a pc-sequence in* K *with pseudolimit* $b \in K$. *Let* $\boldsymbol{b} := c_K(b_\rho)$ *and*

$$\mathcal{E}_{\boldsymbol{b}} := \{g \in K^\times : g \prec b_\rho - b, \text{ eventually}\}.$$

(i) *If* $\mathcal{E}_{\boldsymbol{b}} \neq \varnothing$, *then* $\operatorname{ddeg}_{\boldsymbol{b}} P = \operatorname{ddeg}_{\mathcal{E}_{\boldsymbol{b}}} P_{+b}$.

(ii) *If* $\mathcal{E}_{\boldsymbol{b}} = \varnothing$, *then* $\operatorname{ddeg}_{\boldsymbol{b}} P = \operatorname{mul} P_{+b}$.

Proof. This follows by applying Lemma 4.6 to $a_\rho := b_\rho - b$ and P_{+b}, using Lemma 2.10(ii). □

5. Asymptotic differential equations

Assumption. In this section, K has a monomial group $\mathfrak{M}$ and Γ has no least positive element.

Let $\mathfrak{m}$ range over $\mathfrak{M}$ and $P \in K\{Y\}^{\neq}$ have order at most r. An *asymptotic differential equation* over K is something of the form

$$P(Y) = 0, \qquad Y \in \mathcal{E}, \tag{E}$$

where $\mathcal{E} \subseteq K^\times$ is $\preccurlyeq$-closed. That is, it consists of an algebraic differential equation with an asymptotic condition on solutions. If $\mathcal{E} = \{g \in K^\times : g \prec f\}$ for some $f \in K^\times$, then we write $Y \prec f$ for the asymptotic condition instead of $Y \in \mathcal{E}$, and similarly with $\preccurlyeq$.

For the rest of this section, we fix such an asymptotic differential equation (E). Then the *dominant degree of* (E) is defined to be $\operatorname{ddeg}_{\mathcal{E}} P$. A *solution of* (E) is a $y \in \mathcal{E}$ such that $P(y) = 0$. An *approximate solution of* (E) is an approximate zero of P that lies in $\mathcal{E}$, and the *multiplicity* of an approximate solution of (E) is its multiplicity as an approximate zero of P. The following is used frequently and follows from Lemma 4.2.

Corollary 5.1. *Let $y \in \mathcal{E}$. Then*

(i) *y is an approximate solution of* (E) *if and only if* $\operatorname{ddeg}_{\prec y} P_{+y} \geqslant 1$;

(ii) *if y is an approximate solution of* (E), *then its multiplicity is* $\operatorname{ddeg}_{\prec y} P_{+y}$.

An *algebraic starting monomial for* (E) is an algebraic starting monomial for P that lies in $\mathcal{E}$. So if $\operatorname{ddeg}_{\mathcal{E}} P = 0$, then (E) has no algebraic starting monomials. By Proposition 4.4, if Γ is divisible and $\operatorname{mul} P < \operatorname{ddeg}_{\mathcal{E}} P$, then there is an algebraic starting monomial for (E) and $\operatorname{ddeg}_{\mathcal{E}} P = \operatorname{ddeg} P_{\times\mathfrak{e}}$, where $\mathfrak{e}$ is the largest algebraic starting monomial for (E).

It will be important to alter P and $\mathcal{E}$ in certain ways. Namely, let $\mathcal{E}' \subseteq \mathcal{E}$ be $\preccurlyeq$-closed and let $f \in \mathcal{E} \cup \{0\}$. We call the asymptotic differential equation

$$P_{+f}(Y) = 0, \qquad Y \in \mathcal{E}' \tag{E$'$}$$

a *refinement of* (E). Below, (E′) refers to a refinement of this form. By Lemma 2.6,

$$\operatorname{ddeg}_{\mathcal{E}} P = \operatorname{ddeg}_{\mathcal{E}} P_{+f} \geqslant \operatorname{ddeg}_{\mathcal{E}'} P_{+f},$$

so the dominant degree of (E′) is at most the dominant degree of (E). Note also that if y is a solution of (E′) and $f + y \neq 0$, then $f + y$ is a solution of (E). The same is true with “approximate solution” replacing “solution”, provided that $y \not\sim -f$. The following routine adaptation of [ADH 2017, Lemma 13.8.2] gives a sufficient condition for being an approximate solution.

Lemma 5.2. *Let $f \neq 0$ with $f \succ g$ for all $g \in \mathcal{E}'$, and suppose that*

$$\operatorname{ddeg}_{\mathcal{E}'} P_{+f} = \operatorname{ddeg}_{\mathcal{E}} P \geqslant 1.$$

Then f is an approximate solution of (E).

Proof. We have, using Lemma 2.6 for the equality,

$$\operatorname{ddeg}_{\mathcal{E}'} P_{+f} \leqslant \operatorname{ddeg}_{\prec f} P_{+f} \leqslant \operatorname{ddeg}_{\preccurlyeq f} P_{+f} = \operatorname{ddeg}_{\preccurlyeq f} P \leqslant \operatorname{ddeg}_{\mathcal{E}} P.$$

Hence $\operatorname{ddeg}_{\prec f} P_{+f} = \operatorname{ddeg}_{\mathcal{E}} P \geqslant 1$, so f is an approximate solution of (E). □

Note that by the previous proof, $\operatorname{ddeg}_{\prec f} P_{+f} \leqslant \operatorname{ddeg}_{\mathcal{E}} P$ for all $f \in \mathcal{E}$. The next lemma, corresponding to [ADH 2017, Lemma 13.8.3], relates the strictness of this inequality to approximate solutions.

Lemma 5.3. *Suppose that* $d := \operatorname{ddeg}_{\mathcal{E}} P \geqslant 1$. *Then the following are equivalent*:

(i) $\operatorname{ddeg}_{\prec f} P_{+f} < d$ *for all* $f \in \mathcal{E}$;

(ii) $\operatorname{ddeg}_{\prec f} P_{+f} < d$ *for all* $f \in \mathcal{E}$ *with* $\operatorname{ddeg} P_{\times f} = d$;

(iii) *there is no approximate solution of* (E) *of multiplicity* d.

Proof. The equivalence of (i) and (iii) is given by Corollary 5.1. Now, let $f \in \mathcal{E}$ and suppose that $\operatorname{ddeg} P_{\times f} < d$. Then, using Lemma 2.6 for the first equality,

$$\operatorname{ddeg}_{\prec f} P_{+f} \leqslant \operatorname{ddeg}_{\preccurlyeq f} P_{+f} = \operatorname{ddeg}_{\preccurlyeq f} P = \operatorname{ddeg} P_{\times f} < d.$$

This gives (ii)⟹(i), and the converse is trivial. □

We say that (E) is *unravelled* if $d := \operatorname{ddeg}_{\mathcal{E}} P \geqslant 1$ and the conditions in Lemma 5.3 hold. In particular, if $d \geqslant 1$ and (E) does not have an approximate solution, then (E) is unravelled. And if (E) is unravelled and has an approximate solution, then $d \geqslant 2$ by Lemma 5.3(iii). We now introduce unravellers and partial unravellers, which correspond to special refinements of (E). In the proof of Proposition 5.5, we construct a sequence of partial unravellers ending in an unravelled asymptotic differential equation. Suppose that $d \geqslant 1$, and let $f \in \mathcal{E} \cup \{0\}$ and $\mathcal{E}' \subseteq \mathcal{E}$ be $\preccurlyeq$-closed. We say that $(f, \mathcal{E}')$ is a *partial unraveller for* (E) if $\operatorname{ddeg}_{\mathcal{E}'} P_{+f} = d$. By Lemma 2.6, $(f, \mathcal{E})$ is a partial unraveller for (E). Note that if $(f, \mathcal{E}')$ is a partial unraveller for (E) and $(f_1, \mathcal{E}_1)$ is a partial unraveller for (E′), then $(f + f_1, \mathcal{E}_1)$ is a partial unraveller for (E). An *unraveller for* (E) is a partial unraveller $(f, \mathcal{E}')$ for (E) with unravelled (E′). The following is routine, corresponding to [ADH 2017, Lemma 13.8.6].

Lemma 5.4. *Suppose that* $\operatorname{ddeg}_{\mathcal{E}} P \geqslant 1$. *Let* $a \in K^\times$ *and set* $a\mathcal{E} := \{ay \in K^\times : y \in \mathcal{E}\}$. *Consider the asymptotic differential equation*

$$(a\mathrm{E}) \qquad P_{\times a^{-1}}(Y) = 0, \qquad Y \in a\mathcal{E}.$$

(i) *The dominant degree of* (aE) *equals the dominant degree of* (E).

(ii) *If* $(f, \mathcal{E}')$ *is a partial unraveller for* (E), *then* $(af, a\mathcal{E}')$ *is a partial unraveller for* (aE).

(iii) *If* $(f, \mathcal{E}')$ *is an unraveller for* (E), *then* $(af, a\mathcal{E}')$ *is an unraveller for* (aE).

(iv) *If $a \in \mathfrak{M}$, then the algebraic starting monomials for* (aE) *are exactly the* $a\mathfrak{e}$, *where* $\mathfrak{e}$ *ranges over the algebraic starting monomials for* (E).

The next proposition is about the existence of unravellers, and is a key ingredient in the proof of Proposition 3.1. It corresponds to [ADH 2017, Proposition 13.8.8], and the main difference in the proof is to invoke r-d-algebraic maximality and the nontriviality of the derivation on $\boldsymbol{k}$ instead of asymptotic d-algebraic maximality to obtain pseudolimits of appropriate pc-sequences. For this, recall the notion of r-d-algebraic maximality from Section 3.C, and in particular that if the derivation induced on $\boldsymbol{k}$ is nontrivial, then K is r-d-algebraically maximal if and only if every pc-sequence in K with minimal d-polynomial over K of order at most r has a pseudolimit in K.

Proposition 5.5. *Suppose that K is r-d-algebraically maximal, Γ is divisible, and the derivation induced on $\boldsymbol{k}$ is nontrivial. Suppose that $d := \operatorname{ddeg}_{\mathcal{E}} P \geqslant 1$ and that there is no $f \in \mathcal{E} \cup \{0\}$ with $\operatorname{mul} P_{+f} = d$. Then there exists an unraveller for* (E).

Proof. We construct a sequence $((f_\lambda, \mathcal{E}_\lambda))_{\lambda<\rho}$ of partial unravellers for (E) indexed by an ordinal $\rho > 0$ such that

(i) $\mathcal{E}_\lambda \supseteq \mathcal{E}_\mu$ for all $\lambda < \mu < \rho$;

(ii) $f_\mu - f_\lambda \succ f_\nu - f_\mu$ for all $\lambda < \mu < \nu < \rho$;

(iii) $f_{\lambda+1} - f_\lambda \in \mathcal{E}_\lambda \setminus \mathcal{E}_{\lambda+1}$ for all λ with $\lambda + 1 < \rho$.

For $\rho = 1$, we set $(f_0, \mathcal{E}_0) := (0, \mathcal{E})$ and these conditions are vacuous. Below, we frequently use that by (ii) we have $f_\mu - f_\lambda \asymp f_{\lambda+1} - f_\lambda$ for all $\lambda < \mu < \rho$.

First, suppose that ρ is a successor ordinal, so $\rho = \sigma + 1$, and consider the refinement

$$(\mathrm{E}_\sigma) \qquad\qquad P_{+f_\sigma}(Y) = 0, \qquad Y \in \mathcal{E}_\sigma$$

of (E). If (E_σ) is unravelled, then $(f_\sigma, \mathcal{E}_\sigma)$ is an unraveller for (E) and we are done, so suppose that (E_σ) is not unravelled. Take $f \in \mathcal{E}_\sigma$ such that $\operatorname{ddeg}_{\prec f}(P_{+f_\sigma})_{+f} = d$. Then

$$\mathcal{E}_\rho := \{y \in K^\times : y \prec f\} \subset \mathcal{E}_\sigma$$

is $\preccurlyeq$-closed with

$$\operatorname{ddeg}_{\mathcal{E}_\rho}(P_{+f_\sigma})_{+f} = d,$$

so $(f, \mathcal{E}_\rho)$ is a partial unraveller for (E_σ). Thus, setting $f_\rho := f_\sigma + f$, we have that $(f_\rho, \mathcal{E}_\rho)$ is a partial unraveller for (E). Conditions (i) and (iii) on $((f_\lambda, \mathcal{E}_\lambda))_{\lambda<\rho+1}$ with $\rho + 1$ in place of ρ are obviously satisfied. For (ii), it is sufficient to check that $f_{\lambda+1} - f_\lambda \succ f_\rho - f_\sigma = f$ for $\lambda < \sigma$, which follows from $f_{\lambda+1} - f_\lambda \notin \mathcal{E}_\sigma$.

Now suppose that ρ is a limit ordinal. By (ii), $(f_\lambda)_{\lambda<\rho}$ is a pc-sequence in K, so we let $\boldsymbol{f} := c_K(f_\lambda)$ and claim that $\operatorname{ddeg}_{\boldsymbol{f}} P = d$. To see this, set $g_\lambda := f_{\lambda+1} - f_\lambda$

for λ with $\lambda+1<\rho$. By (iii), we have, using Lemma 2.6 in the third line,

$$\begin{aligned} d = \operatorname{ddeg}_{\mathcal{E}_{\lambda+1}} P_{+f_{\lambda+1}} &\leqslant \operatorname{ddeg}_{\preccurlyeq g_\lambda} P_{+f_{\lambda+1}} \\ &= \operatorname{ddeg}_{\preccurlyeq g_\lambda} (P_{+f_\lambda})_{+(f_{\lambda+1}-f_\lambda)} \\ &= \operatorname{ddeg}_{\preccurlyeq g_\lambda} P_{+f_\lambda} \\ &\leqslant \operatorname{ddeg}_{\mathcal{E}_\lambda} P_{+f_\lambda} = d. \end{aligned}$$

Thus $\operatorname{ddeg}_{\preccurlyeq g_\lambda} P_{+f_\lambda} = d$ for all $\lambda<\rho$, so $\operatorname{ddeg}_{\boldsymbol{f}} P = d$. By Lemma 2.16 and Corollary 2.17, $(f_\lambda)_{\lambda<\rho}$ has a minimal d-polynomial over K of order at most r, so since K is r-d-algebraically maximal, we may take $f_\rho \in K$ with $f_\lambda \rightsquigarrow f_\rho$. Now set

$$\mathcal{E}_\rho := \bigcap_{\lambda<\rho} \mathcal{E}_\lambda = \{y \in K^\times : y \prec g_\lambda \text{ for all } \lambda<\rho\},$$

where the equality follows from (iii). If $\mathcal{E}_\rho=\varnothing$, then by Corollary 4.7,

$$d = \operatorname{ddeg}_{\boldsymbol{f}} P = \operatorname{mul} P_{+f_\rho},$$

contradicting the hypothesis. So $\mathcal{E}_\rho \neq \varnothing$, and thus Corollary 4.7 yields

$$d = \operatorname{ddeg}_{\boldsymbol{f}} P = \operatorname{ddeg}_{\mathcal{E}_\rho} P_{+f_\rho},$$

so $(f_\rho, \mathcal{E}_\rho)$ is a partial unraveller for (E). For $((f_\lambda, \mathcal{E}_\lambda))_{\lambda<\rho+1}$, conditions (i) and (iii) with $\rho+1$ in place of ρ are obviously satisfied. For (ii), it is enough to check that $f_{\lambda+1}-f_\lambda \succ f_\rho - f_\mu$ for $\lambda<\mu<\rho$, which follows from $f_\rho-f_\mu \asymp f_{\mu+1}-f_\mu$.

This inductive construction must end, and therefore there exists an unraveller for (E). □

5.A. ***Behaviour of unravellers under immediate extensions.*** In this subsection, we fix an immediate extension L of K, and we use the monomial group of K as a monomial group for L. We consider how unravellers change under passing from K to L and connect this to pseudolimits of pc-sequences. Lemma 5.8 is a key step in the proof of Proposition 3.1.

Given $\mathcal{E}$, the set $\mathcal{E}_L := \{y \in L^\times : vy \in v\mathcal{E}\}$ is also $\preccurlyeq$-closed with $\mathcal{E}_L \cap K = \mathcal{E}$. Consider the asymptotic differential equation

$$(\mathrm{E}_L) \qquad P(Y)=0, \qquad Y \in \mathcal{E}_L$$

over L, which has the same dominant degree as (E), i.e., $\operatorname{ddeg}_{\mathcal{E}_L} P = \operatorname{ddeg}_{\mathcal{E}} P$. Note that $y \in K$ is an approximate solution of (E) if and only if it is an approximate solution of (E_L). If so, its multiplicities in both settings agree. Thus if (E_L) is unravelled, then (E) is unravelled. For the other direction, if $y \in L$ is an approximate solution of (E_L) of multiplicity $\operatorname{ddeg}_{\mathcal{E}_L} P$, then any $z \in K$ with $z \sim y$ is an approximate solution of (E) of multiplicity $\operatorname{ddeg}_{\mathcal{E}} P = \operatorname{ddeg}_{\mathcal{E}_L} P$. The next lemma follows from this, and corresponds to [ADH 2017, Lemma 13.8.9].

Lemma 5.6. *Suppose that* $\operatorname{ddeg}_{\mathcal{E}} P \geqslant 1$, *and let* $f \in \mathcal{E} \cup \{0\}$ *and* $\mathcal{E}' \subseteq \mathcal{E}$ *be* $\preccurlyeq$*-closed. Then*

(i) $(f, \mathcal{E}')$ *is a partial unraveller for* (E) *if and only if* $(f, \mathcal{E}'_L)$ *is a partial unraveller for* (E_L);

(ii) $(f, \mathcal{E}')$ *is an unraveller for* (E) *if and only if* $(f, \mathcal{E}'_L)$ *is an unraveller for* (E_L).

This next lemma does not use the assumptions of this section. It corresponds to [ADH 2017, Lemma 13.8.10] and has exactly the same proof, except for using Lemma 2.11 instead of [ADH 2017, Lemma 11.3.8].

Lemma 5.7. *Suppose that the derivation induced on* $\boldsymbol{k}$ *is nontrivial. Let* (a_ρ) *be a divergent pc-sequence in* K *with minimal d-polynomial* G *over* K, *and let* $a_\rho \rightsquigarrow \ell \in L$. *Then* $\operatorname{mul}(G_{+\ell}) \leqslant 1$.

The next lemma is routinely adapted from [ADH 2017, Lemma 13.8.11].

Lemma 5.8. *Suppose that* Γ *is divisible and the derivation induced on* $\boldsymbol{k}$ *is nontrivial. Let* (a_ρ) *be a divergent pc-sequence in* K *with minimal d-polynomial* P *over* K, *and* $a_\rho \rightsquigarrow \ell \in L$. *Suppose that* L *is* r*-d-algebraically maximal and* $\operatorname{ddeg}_{\boldsymbol{a}} P \geqslant 2$. *Let* $a \in K$ *and* $\mathfrak{v} \in K^\times$ *be such that* $a - \ell \prec \mathfrak{v}$ *and* $\operatorname{ddeg}_{\prec \mathfrak{v}} P_{+a} = \operatorname{ddeg}_{\boldsymbol{a}} P$. (Such a and $\mathfrak{v}$ exist by Lemma 2.16.) *Consider the asymptotic differential equation*

$$P_{+a}(Y) = 0, \qquad Y \prec \mathfrak{v}. \tag{5-1}$$

Then there exists an unraveller $(f, \mathcal{E})$ *for* (5-1) *over* L *such that*

(i) $f \neq 0$;

(ii) $\operatorname{ddeg}_{\prec f} P_{+a+f} = \operatorname{ddeg}_{\boldsymbol{a}} P$;

(iii) $a_\rho \rightsquigarrow a + f + z$ *for all* $z \in \mathcal{E} \cup \{0\}$.

Proof. We first show how to arrange that $a = 0$ and (ii) holds. Take $g \in K^\times$ with $a - \ell \sim -g$, so $g \prec \mathfrak{v}$. Then, using Lemma 2.6, we have

$$\operatorname{ddeg}_{\prec g} P_{+a+g} \leqslant \operatorname{ddeg}_{\prec \mathfrak{v}} P_{+a+g} = \operatorname{ddeg}_{\prec \mathfrak{v}} P_{+a} = \operatorname{ddeg}_{\boldsymbol{a}} P.$$

Conversely, as $(a+g) - \ell \prec g$, Lemma 2.16 gives $\operatorname{ddeg}_{\boldsymbol{a}} P \leqslant \operatorname{ddeg}_{\prec g} P_{+a+g}$, so

$$\operatorname{ddeg}_{\boldsymbol{a}} P = \operatorname{ddeg}_{\prec \mathfrak{v}} P_{+a} = \operatorname{ddeg}_{\prec g} P_{+a+g}.$$

P_{+a+g} is a minimal d-polynomial of $(a_\rho - (a+g))$ over K and, by Lemma 2.10(ii),

$$\operatorname{ddeg}_{\boldsymbol{a}-(a+g)} P_{+a+g} = \operatorname{ddeg}_{\boldsymbol{a}} P.$$

We can now replace P, (a_ρ), ℓ, and $\mathfrak{v}$ with P_{+a+g}, $(a_\rho - (a+g))$, $\ell - (a+g)$, and g, respectively. To see that this works, suppose that $\mathcal{E} \subseteq L^\times$ is $\preccurlyeq$-closed in L with $\mathcal{E} \prec g$, and $(h, \mathcal{E})$ is an unraveller for the asymptotic differential equation

$$P_{+a+g}(Y) = 0, \qquad Y \prec g$$

over L with $a_\rho - (a+g) \rightsquigarrow h+z$ for all $z \in \mathcal{E} \cup \{0\}$. In particular, $h \prec g$, so $g+h \neq 0$, and it is clear from $\operatorname{ddeg}_{\prec \mathfrak{v}} P_{+a} = \operatorname{ddeg}_{\prec g} P_{+a+g}$ that $(g+h, \mathcal{E})$ is an unraveller for (5-1). Condition (iii) is also obviously satisfied. For condition (ii), note that as $h \prec g$, using Lemma 2.6 in the middle equality,

$$\operatorname{ddeg}_{\prec g+h} P_{+a+g+h} = \operatorname{ddeg}_{\prec g} P_{+a+g+h} = \operatorname{ddeg}_{\prec g} P_{+a+g} = \operatorname{ddeg}_{\boldsymbol{a}} P.$$

Thus it remains to show that there is an unraveller $(f, \mathcal{E})$ for (5-1) in L (with $a=0$) such that $a_\rho \rightsquigarrow f+z$ for all $z \in \mathcal{E} \cup \{0\}$. Consider the set

$$\mathcal{Z} := \{z \in L^\times : z \prec a_\rho - \ell, \text{ eventually}\}.$$

For any $z \in \mathcal{Z} \cup \{0\}$, we have $a_\rho \rightsquigarrow z+\ell$, so by Lemma 5.7,

$$\operatorname{mul}(P_{+\ell+z}) \leqslant 1 < 2 \leqslant \operatorname{ddeg}_{\boldsymbol{a}} P.$$

By Corollary 4.7, $\mathcal{Z} \neq \emptyset$, so $\mathcal{Z}$ is $\preccurlyeq$-closed and $\operatorname{ddeg}_{\mathcal{Z}} P_{+\ell} = \operatorname{ddeg}_{\boldsymbol{a}} P$. Then Proposition 5.5 yields an unraveller $(s, \mathcal{E})$ for the asymptotic differential equation

$$P_{+\ell}(Y) = 0, \qquad Y \in \mathcal{Z}$$

over L. Setting $f := \ell + s$, we get that $(f, \mathcal{E})$ is an unraveller for (5-1) with $a_\rho \rightsquigarrow f+z$ for all $z \in \mathcal{E} \cup \{0\}$. □

5.B. *Reducing degree.* In this subsection, we consider a refinement of (E) and then truncate it by removing monomials of degree higher than the dominant degree of (E). Given an unraveller for (E), we show how to find an unraveller for this truncated refinement in Lemma 5.10, an essential component in the proof of Proposition 6.1.

Assumption. In this subsection, Γ is divisible.

Suppose that $d := \operatorname{ddeg}_{\mathcal{E}} P \geqslant 1$ and we have an unraveller $(f, \mathcal{E}')$ for (E). That is, the refinement

$$\text{(E}'\text{)} \qquad P_{+f}(Y) = 0, \qquad Y \in \mathcal{E}'$$

of (E) is unravelled with dominant degree d. Now suppose that $d > \operatorname{mul}(P_{+f})$, so (E$'$) has an algebraic starting monomial, and let $\mathfrak{e}$ be its largest algebraic starting monomial. Suppose that $g \in K^\times$ satisfies $\mathfrak{e} \prec g \prec f$, and consider another refinement of (E):

$$\text{(E}_g\text{)} \qquad P_{+f-g}(Y) = 0, \qquad Y \preccurlyeq g.$$

Set $\mathcal{E}'_g := \{y \in \mathcal{E}' : y \prec g\}$, so $\mathfrak{e} \in \mathcal{E}'_g$. The next lemma is routinely adapted from [ADH 2017, Lemma 13.8.12].

Lemma 5.9. *The asymptotic differential equation* (E$_g$) *has dominant degree d and* $(g, \mathcal{E}'_g)$ *is an unraveller for* (E$_g$).

Proof. First, since $\mathfrak{e}$ is the largest algebraic starting monomial for (E′), Proposition 4.4 gives

$$d = \operatorname{ddeg} P_{+f,\times\mathfrak{e}} = \operatorname{ddeg}_{\preccurlyeq\mathfrak{e}} P_{+f}.$$

Note that $f - g \sim f \in \mathcal{E}$. Now, by Lemma 2.6 we obtain

$$d = \operatorname{ddeg}_{\preccurlyeq\mathfrak{e}} P_{+f} \leqslant \operatorname{ddeg}_{\preccurlyeq g} P_{+f} = \operatorname{ddeg}_{\preccurlyeq g} P_{+f-g} \leqslant \operatorname{ddeg}_{\mathcal{E}} P_{+f-g} = \operatorname{ddeg}_{\mathcal{E}} P = d,$$

which gives that (E_g) has dominant degree d. Similarly,

$$d = \operatorname{ddeg}_{\preccurlyeq\mathfrak{e}} P_{+f} \leqslant \operatorname{ddeg}_{\mathcal{E}'_g} P_{+f} \leqslant \operatorname{ddeg}_{\mathcal{E}} P_{+f} = \operatorname{ddeg}_{\mathcal{E}} P = d,$$

and thus the asymptotic differential equation

$$P_{+f}(Y) = 0, \qquad Y \in \mathcal{E}'_g,$$

which is a refinement of both (E_g) and (E′), has dominant degree d. Finally, since (E′) is unravelled, the pair $(g, \mathcal{E}'_g)$ is an unraveller for (E_g). □

We now turn to ignoring terms of degree higher than the dominant degree of (E). First, some notation. Recall that for $F \in K\{Y\}$, we set $F_{\leqslant n} := F_0 + F_1 + \cdots + F_n$. Note that if $n \geqslant \operatorname{ddeg} F$, then $D_F = D_{F_{\leqslant n}}$. Now set $F := P_{+f-g}$, so $d \geqslant \operatorname{ddeg} F_{\times\mathfrak{m}}$ for all $\mathfrak{m} \preccurlyeq g$. Consider the "truncation"

$$(\mathrm{E}_{g,\leqslant d}) \qquad F_{\leqslant d}(Y) = 0, \qquad Y \preccurlyeq g$$

of (E_g) as an asymptotic differential equation over K. We have, for all $\mathfrak{m} \preccurlyeq g$,

$$D_{F_{\times\mathfrak{m}}} = D_{(F_{\times\mathfrak{m}})_{\leqslant d}} = D_{(F_{\leqslant d})_{\times\mathfrak{m}}},$$

so ($\mathrm{E}_{g,\leqslant d}$) has the same algebraic starting monomials and dominant degree as (E_g). Next, we show that under suitable conditions the unraveller $(g, \mathcal{E}'_g)$ for (E_g) from the previous lemma remains an unraveller for ($\mathrm{E}_{g,\leqslant d}$). Recall that $[\gamma]$ denotes the archimedean class of $\gamma \in \Gamma$ and that such classes are ordered in the natural way; see Section 2.E.

The next lemma corresponds to [ADH 2017, Lemma 13.8.13]. The essential difference is that we use the valuation $vg \mapsto [vg]$ on Γ instead of the valuation $vg \mapsto v(g'/g)$ used in Lemma 13.8.13.

Lemma 5.10. *Suppose that* $[v(\mathfrak{e}/g)] < [v(g/f)]$. *Then* $(g, \mathcal{E}'_g)$ *is an unraveller for* ($\mathrm{E}_{g,\leqslant d}$), *and* $\mathfrak{e}$ *is the largest algebraic starting monomial for the unravelled asymptotic differential equation*

$$(\mathrm{E}'_{g,\leqslant d}) \qquad (F_{\leqslant d})_{+g}(Y) = 0, \qquad Y \in \mathcal{E}'_g.$$

Proof. First, we reduce to the case $g \asymp 1$: set $\mathfrak{g} := \mathfrak{d}_g$ and replace P, f, g, $\mathcal{E}$, and $\mathcal{E}'$ by $P_{\times\mathfrak{g}}$, $f/\mathfrak{g}$, $g/\mathfrak{g}$, $\mathfrak{g}^{-1}\mathcal{E}$, and $\mathfrak{g}^{-1}\mathcal{E}'$, respectively, and use Lemma 5.4. Note that now $\mathfrak{e} \prec 1 \prec f$ and $[v\mathfrak{e}] < [vf]$.

Since $F = P_{+f-g}$ and $g \asymp 1$, we have $\operatorname{ddeg} F = \operatorname{ddeg}_{\preccurlyeq 1} F = d$ by Lemma 5.9, so

$$d \leqslant \operatorname{ddeg} F_{\times f} \leqslant \operatorname{ddeg}_{\mathcal{E}} F = d,$$

using Corollary 2.4 and Lemma 2.6. This yields $d = \operatorname{ddeg} F = \operatorname{ddeg} F_{\times f}$. For $\mathfrak{m}$ with $[v\mathfrak{m}] < [vf]$ we may thus apply Corollary 2.27 with F and $\mathfrak{d}_f$ in place of P and $\mathfrak{n}$ to get

$$D_{P_{+f,\times\mathfrak{m}}} = D_{F_{+g,\times\mathfrak{m}}} = D_{(F_{\leqslant d})_{+g,\times\mathfrak{m}}}. \tag{5-2}$$

In particular, this holds if $\mathfrak{e} \preccurlyeq \mathfrak{m} \prec 1$, as then $[v\mathfrak{m}] \leqslant [v\mathfrak{e}] < [vf]$. Thus $\mathfrak{e}$ is the largest algebraic starting monomial for $(\mathrm{E}'_{g,\leqslant d})$, since it is the largest such for (E').

For $(g, \mathcal{E}'_g)$ to be an unraveller for $(\mathrm{E}_{g,\leqslant d})$, we now show

(i) $\operatorname{ddeg}_{\mathcal{E}'_g} (F_{\leqslant d})_{+g} = d$;

(ii) $\operatorname{ddeg}_{\prec h} (F_{\leqslant d})_{+g+h} < d$ for all $h \in \mathcal{E}'_g$.

For (i), if $\mathfrak{e} \preccurlyeq \mathfrak{m} \in \mathcal{E}'_g$, then by Corollary 4.5 and (5-2) we have

$$d = \operatorname{ddeg} P_{+f,\times\mathfrak{m}} = \operatorname{ddeg}(F_{\leqslant d})_{+g,\times\mathfrak{m}}.$$

For (ii), let $h \in \mathcal{E}'_g$, so $h \in \mathcal{E}'$ and $h \prec 1$. Set $\mathfrak{h} := \mathfrak{d}_h$ and $u := h/\mathfrak{h}$. Applying Lemma 4.2, we have

$$\operatorname{ddeg}_{\prec h} (F_{\leqslant d})_{+g+h} = \operatorname{mul}\big(D_{(F_{\leqslant d})_{+g,\times\mathfrak{h}}}\big)_{+\bar{u}}, \tag{5-3}$$

$$\operatorname{ddeg}_{\prec h} P_{+f+h} = \operatorname{mul}\big(D_{P_{+f,\times\mathfrak{h}}}\big)_{+\bar{u}}. \tag{5-4}$$

First suppose $\mathfrak{e} \preccurlyeq h$, so then combining (5-2), for $\mathfrak{m} = \mathfrak{h}$, with (5-3) and (5-4) we have

$$\operatorname{ddeg}_{\prec h} (F_{\leqslant d})_{+g+h} = \operatorname{ddeg}_{\prec h} P_{+f+h} < d,$$

since (E') is unravelled. Now suppose $h \prec \mathfrak{e}$. If $\mathfrak{e}^2 \preccurlyeq h \prec \mathfrak{e}$, then $[vh] = [v\mathfrak{e}] < [vf]$, and thus by (5-2) and Corollary 4.5,

$$\operatorname{ddeg}(F_{\leqslant d})_{+g,\times\mathfrak{h}} = \operatorname{ddeg} P_{+f,\times\mathfrak{h}} < \operatorname{ddeg} P_{+f,\times\mathfrak{e}} = d.$$

By Corollary 2.4, $\operatorname{ddeg}(F_{\leqslant d})_{+g,\times\mathfrak{h}} < d$ remains true for any $h \prec \mathfrak{e}$. Hence, by (5-3),

$$\operatorname{ddeg}_{\prec h} (F_{\leqslant d})_{+g+h} = \operatorname{mul}\big(D_{(F_{\leqslant d})_{+g,\times\mathfrak{h}}}\big)_{+\bar{u}} \leqslant \operatorname{ddeg}(F_{\leqslant d})_{+g,\times\mathfrak{h}} < d,$$

which completes the proof of (ii). □

5.C. ***Finding solutions in differential-henselian fields.*** We now use d-henselianity to find solutions of asymptotic differential equations. Given an element of an extension of K, when K has few constants we find a solution closest to that element. The only result in this subsection that uses the assumption that Γ has no least positive element is Lemma 5.15.

We say that (E) is *quasilinear* if $\operatorname{ddeg}_{\mathcal{E}} P = 1$. Note that if K is r-d-henselian and (E) is quasilinear, then P has a zero in $\mathcal{E} \cup \{0\}$. Note that even in this case, (E) may not have a solution, since those are required to be nonzero. The next lemma, routinely adapted from [ADH 2017, Lemma 14.3.4], shows how certain approximate solutions yield solutions.

Lemma 5.11. *Suppose that K is r-d-henselian. Let $g \in K^\times$ be an approximate zero of P such that* $\operatorname{ddeg} P_{\times g} = 1$. *Then there exists $y \sim g$ in K such that $P(y) = 0$.*

Proof. Let $\mathfrak{m} := \mathfrak{d}_g$ and $u := g/\mathfrak{m}$, so $D_{P_{\times \mathfrak{m}}}(\bar{u}) = 0$ and thus

$$\operatorname{dmul} P_{\times \mathfrak{m}, +u} = \operatorname{mul}(D_{P_{\times \mathfrak{m}}})_{+\bar{u}} \geqslant 1.$$

By Lemma 2.2, we also have

$$\operatorname{dmul} P_{\times \mathfrak{m}, +u} \leqslant \operatorname{ddeg} P_{\times \mathfrak{m}, +u} = \operatorname{ddeg} P_{\times \mathfrak{m}} = 1.$$

Thus $\operatorname{dmul} P_{\times \mathfrak{m}, +u} = 1$, so by r-d-henselianity we have $z \prec 1$ with $P_{\times \mathfrak{m}, +u}(z) = 0$. Setting $y := (u + z)\mathfrak{m}$ gives $P(y) = 0$ and $y \sim u\mathfrak{m} = g$. □

Now let f be an element of an extension of K. We say that a solution y of (E) *best approximates f (among solutions of* (E)) if $y - f \preccurlyeq z - f$ for each solution z of (E). Note that if $f \in K^\times$ is a solution of (E), then f is the unique solution of (E) that best approximates f. Also, if $f \succ \mathcal{E}$, then $y - f \asymp f$ for all $y \in \mathcal{E}$, and so each solution of (E) best approximates f. The next lemma concerning multiplicative conjugation has a routine proof, identical to that of [ADH 2017, Lemma 11.2.10], by distinguishing the cases $z \succ g$ and $z \preccurlyeq g$.

Lemma 5.12. *Let f be an element of an extension of K and let $g \in \mathcal{E}$ with $f \preccurlyeq g$. Suppose that y is a solution of the asymptotic differential equation*

$$P_{\times g}(Y) = 0, \qquad Y \preccurlyeq 1$$

that best approximates $g^{-1} f$. Then the solution gy of (E) *best approximates f.*

This lemma is easily adapted from [ADH 2017, Lemma 14.1.13]; it is the one place in this section that we impose the assumption $C \subseteq \mathcal{O}$.

Lemma 5.13. *Suppose that $r \geqslant 1$ and K is r-d-henselian with $C \subseteq \mathcal{O}$. Suppose that* (E) *is quasilinear and has a solution. Let f be an element of an extension of K. Then f is best approximated by some solution of* (E).

Proof. By the comment above Lemma 5.12, we may assume that $f \not\succ \mathcal{E}$. Thus we may take $g \in \mathcal{E}$ with $f \preccurlyeq g$ such that (E) has a solution $y \preccurlyeq g$ and

$$\operatorname{ddeg} P_{\times g} = \operatorname{ddeg}_{\mathcal{E}} P = 1.$$

By Lemma 5.12, we may replace P by $P_{\times g}$ and $\mathcal{E}$ by $\mathcal{O}^{\neq}$ in order to assume that $\mathcal{E} = \mathcal{O}^{\neq}$. Suppose that f is not best approximated by any solution of (E). Then for each i we get $y_i \in K^\times$ such that

(i) y_i is a solution of (E), i.e., $P(y_i) = 0$ and $y_i \preccurlyeq 1$;

(ii) $y_i - f \succ y_{i+1} - f$;

(iii) $\operatorname{ddeg} P_{+y_i} = \operatorname{ddeg} P = 1$ (by Lemma 2.2).

Item (ii) implies $y_{i+1} - y_i \asymp y_i - f$, contradicting [ADH 2017, Lemma 7.5.5]. □

The next lemma is based on [ADH 2017, Lemma 14.1.14] but has a shorter proof using Lemma 5.11.

Lemma 5.14. *Suppose that K is r-d-henselian,* (E) *is quasilinear, and $f \in \mathcal{E}$ is an approximate solution of* (E). *Then* (E) *has a solution $y_0 \sim f$, and every solution y of* (E) *that best approximates f satisfies $y \sim f$.*

Proof. Let $\mathfrak{m} := \mathfrak{d}_f$ and $u := f/\mathfrak{m}$. Then by Lemma 2.2 we have

$$\operatorname{ddeg} P_{\times f} = \operatorname{ddeg} P_{\times \mathfrak{m}, +u} \geqslant \operatorname{dmul} P_{\times \mathfrak{m}, +u} \geqslant 1,$$

and so since (E) is quasilinear, $\operatorname{ddeg} P_{\times f} = 1$. Thus Lemma 5.11 yields a solution $y_0 \sim f$ of (E). If y is a solution of (E) that best approximates f, then $y \sim f$, as

$$y - f \preccurlyeq y_0 - f \prec f. \qquad \square$$

For the next lemma, a routine adaptation of [ADH 2017, Lemma 14.3.13], recall from Section 5.A that given an immediate extension L of K, we extend the asymptotic differential equation (E) over K to (E_L) over L. Note that if (E) is quasilinear, then so is (E_L).

Lemma 5.15. *Suppose that K is r-d-henselian and let L be an immediate extension of K. Suppose that* (E) *is quasilinear, $\mathcal{E}' \subseteq \mathcal{E}$ is $\preccurlyeq$-closed, and $f \in \mathcal{E}_L$ is such that the refinement*

$$(\mathrm{E}'_L) \qquad P_{+f}(Y) = 0, \qquad Y \in \mathcal{E}'_L$$

of (E_L) *is also quasilinear. Let $y \preccurlyeq f$ be a solution of* (E) *that best approximates f. Then $f - y \in \mathcal{E}'_L \cup \{0\}$.*

Proof. The case $f = y$ being trivial, suppose that $f \neq y$ and set $\mathfrak{m} := \mathfrak{d}_{f-y}$. As $f - y \in \mathcal{E}_L$, we have $\mathfrak{m} \in \mathcal{E}$. Now suppose towards a contradiction that $f - y \notin \mathcal{E}'_L$. Then $\mathcal{E}'_L \prec \mathfrak{m} \in \mathcal{E}$, so by quasilinearity and Lemma 2.6,

$$1 = \operatorname{ddeg}_{\mathcal{E}'_L} P_{+f} \leqslant \operatorname{ddeg}_{\preccurlyeq \mathfrak{m}} P_{+f} = \operatorname{ddeg}_{\preccurlyeq \mathfrak{m}} P_{+y} \leqslant \operatorname{ddeg}_{\mathcal{E}} P_{+y} = \operatorname{ddeg}_{\mathcal{E}} P = 1.$$

Hence the asymptotic differential equation

$$(5\text{-}5) \qquad P_{+y}(Y) = 0, \qquad Y \preccurlyeq \mathfrak{m}$$

over K is also quasilinear. Also, by the quasilinearity of (E'_L), we have

$$\operatorname{ddeg}_{\prec\mathfrak{m}}(P_{+y})_{+(f-y)} = \operatorname{ddeg}_{\prec\mathfrak{m}} P_{+f} \geqslant \operatorname{ddeg}_{\mathcal{E}'_L} P_{+f} = 1,$$

so $f - y$ is an approximate solution of (5-5) over L, by Corollary 5.1. Take $g \in K^\times$ with $g \sim f - y$, so g is an approximate solution of (5-5) over K, and, by the quasilinearity of (5-5),

$$\operatorname{ddeg} P_{+y,\times g} = \operatorname{ddeg}_{\preccurlyeq\mathfrak{m}} P_{+y} = 1.$$

Then by Lemma 5.11 we have $z \sim g \sim f - y$ in K such that $P(y+z) = 0$. We must have $y + z \neq 0$, as otherwise $f \prec y - f$, contradicting $y \preccurlyeq f$. From $y \preccurlyeq f$, we also obtain $y + z \preccurlyeq f$, so $y + z \in \mathcal{E}$. Since $y + z - f \prec y - f$, this contradicts that y best approximates f. □

6. Reducing complexity

This is a technical section whose main goal is Proposition 6.1. This proposition, or rather its consequence Corollary 6.14, is the linchpin of Proposition 3.1, and its proof uses all of the previous sections and some additional results from [ADH 2017]. This section is based on [ADH 2017, §14.4].

Assumption. In this section, K is asymptotic and has a monomial group $\mathfrak{M}$, Γ is divisible, and $\boldsymbol{k}$ is r-linearly surjective with $r \geqslant 1$.

Let $\mathfrak{m}$ and $\mathfrak{n}$ range over $\mathfrak{M}$. As usual, we let $P \in K\{Y\}^{\neq}$ with order at most r. As in the previous section, let $\mathcal{E} \subseteq K^\times$ be $\preccurlyeq$-closed, so we have an asymptotic differential equation

$$\text{(E)} \qquad P(Y) = 0, \qquad Y \in \mathcal{E}$$

over K. Set $d := \operatorname{ddeg}_{\mathcal{E}} P$ and suppose that $d \geqslant 2$. We fix an immediate asymptotic r-d-henselian extension $\hat{K}$ of K and use $\mathfrak{M}$ as a monomial group of $\hat{K}$.

Let $\hat{\mathcal{E}} := \mathcal{E}_{\hat{K}} = \{y \in \hat{K}^\times : vy \in v\mathcal{E}\}$, so we have the asymptotic differential equation

$$(\hat{\mathrm{E}}) \qquad P(Y) = 0, \qquad Y \in \hat{\mathcal{E}}$$

over $\hat{K}$ with dominant degree d. Suppose that ($\hat{\mathrm{E}}$) is not unravelled, and that this is witnessed by an $\hat{f} \in \hat{\mathcal{E}}$ such that $(\hat{f}, \hat{\mathcal{E}}')$ is an unraveller for ($\hat{\mathrm{E}}$). That is, $\operatorname{ddeg}_{\prec\hat{f}} P_{+\hat{f}} = d$, and the refinement

$$(\hat{\mathrm{E}}') \qquad P_{+\hat{f}}(Y) = 0, \qquad Y \in \hat{\mathcal{E}}'$$

of ($\hat{\mathrm{E}}$) is unravelled with dominant degree d. By Corollary 5.1, $\hat{f}$ is an approximate solution of ($\hat{\mathrm{E}}$) of multiplicity d. Note also that $\hat{\mathcal{E}}' = \mathcal{E}'_{\hat{K}}$ for the $\preccurlyeq$-closed set $\mathcal{E}' := \hat{\mathcal{E}}' \cap K \subseteq \mathcal{E}$. Since ($\hat{\mathrm{E}}$) is not unravelled, neither is (E) by the discussion preceding Lemma 5.6. Suppose also that mul $P_{+\hat{f}} < d$, so that by Proposition 4.4, ($\hat{\mathrm{E}}'$) has an algebraic starting monomial; let $\mathfrak{e}$ be the largest such.

The main proposition corresponds to [ADH 2017, Proposition 14.4.1].

Proposition 6.1. *There exists $f \in \hat{K}$ such that one of the following holds*:

(i) $\hat{f} - f \preccurlyeq \mathfrak{e}$ *and* $A(f) = 0$ *for some* $A \in K\{Y\}$ *with* $c(A) < c(P)$ *and* $\deg A = 1$;

(ii) $\hat{f} \sim f$, $\hat{f} - a \preccurlyeq f - a$ *for all* $a \in K$, *and* $A(f) = 0$ *for some* $A \in K\{Y\}$ *with* $c(A) < c(P)$ *and* $\operatorname{ddeg} A_{\times f} = 1$.

6.A. ***Special case.*** We first prove Proposition 6.1 in the special case that $\operatorname{ddeg}_{\mathcal{E}} P = \deg P$ and later reduce to this case using Lemma 6.13. Below, we consider the d-polynomial

$$P_{+\hat{f},\times\mathfrak{e}} \in \hat{K}\{Y\};$$

note that $\operatorname{ddeg} P_{+\hat{f},\times\mathfrak{e}} = d$ by the choice of $\mathfrak{e}$. Let $s \leqslant r$ be the order of P. For $\boldsymbol{i} \in \mathbb{N}^{1+s}$, we let

$$\partial^{\boldsymbol{i}} := \frac{\partial^{|\boldsymbol{i}|}}{\partial Y^{i_0} \cdots \partial (Y^{(s)})^{i_s}}$$

denote the partial differential operator on $\hat{K}\{Y\}$ that differentiates i_n times with respect to $Y^{(n)}$ for $n = 0, \ldots, s$. (We also use additive and multiplicative conjugates of partial differential operators; see [ADH 2017, §12.8].) For any partial differential operator (in the sense of [ADH 2017, §12.7]) Δ on $\hat{K}\{Y\}$, any $Q \in \hat{K}\{Y\}$, and any $a \in \hat{K}$,

$$\Delta(Q_{+a}) = (\Delta Q)_{+a}$$

by [ADH 2017, Lemma 12.8.7], so we write ΔQ_{+a} and do not distinguish between these. If $a \in \hat{K}^{\times}$, note that, by [ADH 2017, Lemma 12.8.8],

$$\Delta Q_{\times a} := \Delta(Q_{\times a}) = (\Delta_{\times a} Q)_{\times a},$$

Note that, when no parentheses are used, we intend additive and multiplicative conjugation of Q to take place before Δ is applied, in order to simplify notation.

Now, choose $\boldsymbol{i} \in \mathbb{N}^{1+s}$ such that $\deg(\partial^{\boldsymbol{i}} Y^{\boldsymbol{j}}) = 1$ for some $\boldsymbol{j} \in \mathbb{N}^{1+s}$ with $|\boldsymbol{j}| = d$ and

$$(P_{+\hat{f},\times\mathfrak{e}})_{\boldsymbol{j}} \asymp P_{+\hat{f},\times\mathfrak{e}}.$$

In particular, $|\boldsymbol{i}| = d - 1$ and

$$\operatorname{ddeg} \partial^{\boldsymbol{i}} P_{+\hat{f},\times\mathfrak{e}} = \deg D_{\partial^{\boldsymbol{i}} P_{+\hat{f},\times\mathfrak{e}}} = \deg \partial^{\boldsymbol{i}} D_{P_{+\hat{f},\times\mathfrak{e}}} = 1.$$

We consider the partial differential operator $\Delta := (\partial^{\boldsymbol{i}})_{\times\mathfrak{e}}$ on $\hat{K}\{Y\}$. We have

$$(\Delta P)_{+\hat{f},\times\mathfrak{e}} = \partial^{\boldsymbol{i}} P_{+\hat{f},\times\mathfrak{e}}$$

by [ADH 2017, Lemmas 12.8.7 and 12.8.8]. Hence the asymptotic differential equation

$$\Delta P_{+\hat{f}}(Y) = 0, \qquad Y \preccurlyeq \mathfrak{e}$$

is quasilinear. In [ADH 2017, §14.4], partial differentiation is also used to obtain a quasilinear asymptotic differential equation. Under the powerful assumption of ω-freeness made in that setting, newton polynomials have a very special form, and so a specific choice of Δ was needed. Here, and in the next subsection, a similar technique works despite the lack of restrictions on dominant parts.

This lemma is routinely adapted from [ADH 2017, Lemma 14.4.2].

Lemma 6.2. *Suppose that* $\mathfrak{e} \prec \hat{f}$ *and the asymptotic differential equation*

$$\Delta P(Y) = 0, \qquad Y \in \hat{\mathcal{E}} \tag{6-1}$$

over $\hat{K}$ *is quasilinear. Then* (6-1) *has a solution* $y \sim \hat{f}$, *and if* f *is any solution of* (6-1) *that best approximates* $\hat{f}$, *then* $f - \hat{f} \preccurlyeq \mathfrak{e}$.

Proof. By Lemma 2.6, we have

$$\operatorname{ddeg}_{\prec \hat{f}} \Delta P_{+\hat{f}} \leqslant \operatorname{ddeg}_{\hat{\mathcal{E}}} \Delta P_{+\hat{f}} = \operatorname{ddeg}_{\hat{\mathcal{E}}} \Delta P = 1.$$

But from $\mathfrak{e} \prec \hat{f}$, we also have

$$1 = \operatorname{ddeg}_{\preccurlyeq \mathfrak{e}} \Delta P_{+\hat{f}} \leqslant \operatorname{ddeg}_{\prec \hat{f}} \Delta P_{+\hat{f}},$$

so $\operatorname{ddeg}_{\prec \hat{f}} \Delta P_{+\hat{f}} = 1$. Then since $\hat{K}$ is r-d-henselian, we get $y \sim \hat{f}$ with $\Delta P(y) = 0$.

For the second statement, the refinement

$$\Delta P_{+\hat{f}}(Y) = 0, \qquad Y \preccurlyeq \mathfrak{e} \tag{6-2}$$

of (6-1) is quasilinear, so we can apply Lemma 5.15 with $\hat{K}$ in the roles of both L and K, and ΔP, $\hat{f}$, f, (6-1), and (6-2) in the roles of P, f, y, (E), and (E'_L), respectively. □

We now conclude the proof of Proposition 6.1 in the case that $\deg P = d$, easily adapted from [ADH 2017, Corollary 14.4.3]. Recall that $d \geqslant 2$.

Lemma 6.3. *Suppose that* $\deg P = d$. *Then there exist* $f \in \hat{K}$ *and* $A \in K\{Y\}$ *such that* $\hat{f} - f \preccurlyeq \mathfrak{e}$, $A(f) = 0$, $c(A) < c(P)$, *and* $\deg A = 1$.

Proof. Since $\deg P = d$, we also have $\deg P_{+\hat{f}, \times \mathfrak{e}} = d$, and hence

$$\deg \Delta P = \deg(\Delta P)_{+\hat{f}, \times \mathfrak{e}} = \deg \partial^{\boldsymbol{i}} P_{+\hat{f}, \times \mathfrak{e}} = 1,$$

by the choice of $\boldsymbol{i}$. Hence (6-1) is quasilinear.

If $\hat{f} \preccurlyeq \mathfrak{e}$, then $f := 0$ and $A := Y$ work, so assume that $\mathfrak{e} \prec \hat{f}$. First, Lemma 6.2 yields a solution $y \sim \hat{f}$ of (6-1). As $\hat{K}$ has few constants, Lemma 5.13 gives that $\hat{f}$

is best approximated by some solution f of (6-1). So applying Lemma 6.2 again, we have $f - \hat{f} \preccurlyeq \mathfrak{e}$. Then $\Delta P(f) = 0$, $c(\Delta P) < c(P)$, and $\deg \Delta P = 1$, so we may take $A := \Delta P$. □

6.B. ***Tschirnhaus refinements.*** Set $\mathfrak{f} := \mathfrak{d}_{\hat{f}}$, and we now consider the d-polynomial $P_{\times\mathfrak{f}} \in K\{Y\}^{\neq}$. If $\mathfrak{e} \succcurlyeq \mathfrak{f}$, then the first case of Proposition 6.1 holds for $f := 0$ and $A := Y$, so in the rest of this subsection and in Sections 6.C and 6.D, we suppose that $\mathfrak{e} \prec \mathfrak{f}$. Then we have, by the choice of $\mathfrak{e}$ and Lemma 2.6,

$$d = \operatorname{ddeg}_{\preccurlyeq\mathfrak{e}} P_{+\hat{f}} \leqslant \operatorname{ddeg}_{\preccurlyeq\mathfrak{f}} P_{+\hat{f}} = \operatorname{ddeg}_{\preccurlyeq\mathfrak{f}} P \leqslant \operatorname{ddeg}_{\hat{\mathcal{E}}} P = d,$$

and thus $\operatorname{ddeg} P_{\times\mathfrak{f}} = d$.

Now, choose $\boldsymbol{i} \in \mathbb{N}^{1+s}$ so that $\deg(\partial^{\boldsymbol{i}} Y^{\boldsymbol{j}})$=1 for some $\boldsymbol{j} \in \mathbb{N}^{1+s}$ with $|\boldsymbol{j}| = d$ and $(P_{\times\mathfrak{f}})_{\boldsymbol{j}} \asymp P_{\times\mathfrak{f}}$. Thus we have $|\boldsymbol{i}| = d - 1$ and

$$D_{\partial^{\boldsymbol{i}} P_{\times\mathfrak{f}}} = \partial^{\boldsymbol{i}} D_{P_{\times\mathfrak{f}}},$$

and so $\operatorname{ddeg} \partial^{\boldsymbol{i}} P_{\times\mathfrak{f}} = 1$. We consider the partial differential operator $\Delta := (\partial^{\boldsymbol{i}})_{\times\mathfrak{f}}$ on $\hat{K}\{Y\}$. By [ADH 2017, Lemma 12.8.8],

$$(\Delta P)_{\times\mathfrak{f}} = \partial^{\boldsymbol{i}} P_{\times\mathfrak{f}},$$

and thus the asymptotic differential equation

$$\Delta P(Y) = 0, \qquad Y \preccurlyeq \mathfrak{f} \tag{6-3}$$

over $\hat{K}$ is quasilinear. The comments from Section 6.A about the difference between the partial differentiation used here and that used in [ADH 2017, §14.4] apply in this subsection as well, and necessitate a slight weakening of the following lemma from its counterpart [ADH 2017, Lemma 14.4.4]. It follows immediately from the quasilinearity of (6-3) by Corollary 2.7.

Lemma 6.4. *Suppose that $f \in \hat{K}$ is a solution of* (6-3). *Then for all $g \in \hat{K}^{\times}$ with $g \preccurlyeq \mathfrak{f}$ we have*

$$\operatorname{mul}(\Delta P)_{+f,\times g} = \operatorname{ddeg}(\Delta P)_{+f,\times g} = 1,$$

and hence $(\Delta P)_{+f}$ has no algebraic starting monomial $\mathfrak{g} \in \mathfrak{M}$ with $\mathfrak{g} \preccurlyeq \mathfrak{f}$.

The next statement is based on [ADH 2017, Lemma 14.4.5] but has a different proof due to the different choices of Δ needed here and in [ADH 2017, §14.4].

Lemma 6.5. *The element $\hat{f} \in \hat{K}$ is an approximate solution of* (6-3).

Proof. Set $u := \hat{f}/\mathfrak{f}$. Since $\hat{f}$ is an approximate zero of P of multiplicity $d = \operatorname{ddeg} P_{\times\mathfrak{f}} = \deg D_{P_{\times\mathfrak{f}}}$,

$$(D_{P_{\times\mathfrak{f}}})_{+\bar{u}} = \sum_{|\boldsymbol{j}|=d} (D_{P_{\times\mathfrak{f}}})_{\boldsymbol{j}} Y^{\boldsymbol{j}},$$

by [ADH 2017, Lemma 4.3.1], where $\boldsymbol{j}$ ranges over $\mathbb{N}^{1+s}$. Then

$$(\partial^{\boldsymbol{i}} D_{P_{\times\mathfrak{f}}})_{+\bar{u}} = \partial^{\boldsymbol{i}} (D_{P_{\times\mathfrak{f}}})_{+\bar{u}} = \sum_{|\boldsymbol{j}|=d} (D_{P_{\times\mathfrak{f}}})_{\boldsymbol{j}} \partial^{\boldsymbol{i}} Y^{\boldsymbol{j}},$$

so the multiplicity of $\partial^{\boldsymbol{i}} D_{P_{\times\mathfrak{f}}}$ at $\bar{u}$ is 1 by the choice of $\boldsymbol{i}$. In view of

$$D_{(\Delta P)_{\times\mathfrak{f}}} = D_{\partial^{\boldsymbol{i}} P_{\times\mathfrak{f}}} = \partial^{\boldsymbol{i}} D_{P_{\times\mathfrak{f}}},$$

$\hat{f}$ is an approximate solution of (6-3). □

Let $f \in \hat{K}$ with $f \sim \hat{f}$, so $\operatorname{ddeg}_{\prec\mathfrak{f}} P_{+f} = \operatorname{ddeg}_{\prec\mathfrak{f}} P_{+\hat{f}} = d$ by Lemma 2.6. That is, the refinement

$$\text{(T)} \qquad P_{+f}(Y) = 0, \qquad Y \prec \mathfrak{f}$$

of (Ê) still has dominant degree d. As $\hat{f}$ is an approximate solution of (6-3), Lemmas 5.14 and 5.13 give a solution $f_0 \in \hat{K}$ of (6-3) that best approximates $\hat{f}$ with $f_0 \sim \hat{f} \sim f$. Thus

$$\operatorname{ddeg}_{\prec\mathfrak{f}} \Delta P_{+f} = \operatorname{ddeg}_{\prec\mathfrak{f}} \Delta P_{+f_0} = 1$$

by Lemmas 2.6 and 6.4. Hence the refinement

$$(\Delta\text{T}) \qquad \Delta P_{+f}(Y) = 0, \qquad Y \prec \mathfrak{f}$$

of (6-3) is also quasilinear.

Definition. A *Tschirnhaus refinement of* (Ê) is an asymptotic differential equation (T) over $\hat{K}$ as above with $\hat{f} \sim f \in \hat{K}$ such that some solution $f_0 \in \hat{K}$ of (6-3) over $\hat{K}$ best approximates $\hat{f}$ and satisfies $f_0 - \hat{f} \sim f - \hat{f}$.

Definition. Let $f, \hat{g} \in \hat{K}$ and $\mathfrak{m}$ satisfy

$$\mathfrak{m} \prec f - \hat{f} \preccurlyeq \hat{g} \prec \mathfrak{f},$$

so in particular $f \sim \hat{f}$. With (T) as above, but not necessarily a Tschirnhaus refinement of (Ê), we say that the refinement

$$\text{(TC)} \qquad P_{+f+\hat{g}}(Y) = 0, \qquad Y \preccurlyeq \mathfrak{m}$$

of (T) is *compatible with* (T) if it has dominant degree d and $\hat{g}$ is not an approximate solution of (ΔT).

The next two lemmas are routine adaptations of [ADH 2017, Lemmas 14.4.7 and 14.4.8].

Lemma 6.6. *Let* $f, f_0, \hat{g} \in \hat{K}$ *and* $\mathfrak{m}$ *be such that*

$$\mathfrak{m} \prec f_0 - \hat{f} \sim f - \hat{f} \preccurlyeq \hat{g} \prec \mathfrak{f},$$

and (TC) *has dominant degree d. Then $\hat{g}$ is an approximate solution of* (T) *and of*

$$(\mathrm{T}_0) \qquad P_{+f_0}(Y) = 0, \qquad Y \prec \mathfrak{f}.$$

Proof. First, $\hat{g}$ is an approximate solution of (T) by Lemma 5.2, since $\mathfrak{m} \prec \hat{g}$ and

$$\operatorname{ddeg}_{\preccurlyeq \mathfrak{m}} P_{+f+\hat{g}} = d = \operatorname{ddeg}_{\prec \mathfrak{f}} P_{+f}.$$

From $f_0 - f \prec f - \hat{f} \preccurlyeq \hat{g}$ and $\mathfrak{m} \prec \hat{g}$, we obtain, using Lemma 2.6 in the first equality,

$$\operatorname{ddeg}_{\prec \hat{g}} P_{+f_0+\hat{g}} = \operatorname{ddeg}_{\prec \hat{g}} P_{+f+\hat{g}} \geqslant \operatorname{ddeg}_{\preccurlyeq \mathfrak{m}} P_{+f+\hat{g}} = d \geqslant 1,$$

so $\hat{g}$ is an approximate solution of (T_0) by Corollary 5.1. □

Lemma 6.7. *Let $f, f_0, \hat{g} \in \hat{K}$ with*

$$f_0 - \hat{f} \sim f - \hat{f} \preccurlyeq \hat{g} \prec \mathfrak{f}.$$

Then $\hat{g}$ is an approximate solution of (ΔT) *if and only if $\hat{g}$ is an approximate solution of*

$$(\Delta\mathrm{T}_0) \qquad \Delta P_{+f_0}(Y) = 0, \qquad Y \prec \mathfrak{f}.$$

Proof. Again, since $f_0 - f \prec \hat{f} - f \preccurlyeq \hat{g}$, by Lemma 2.6 we have

$$\operatorname{ddeg}_{\prec \hat{g}} \Delta P_{+f_0+\hat{g}} = \operatorname{ddeg}_{\prec \hat{g}} \Delta P_{+f+\hat{g}}.$$

The result then follows from Corollary 5.1, since $\hat{g} \prec \mathfrak{f}$. □

Note that, for any $f_0 \sim f$, the equation ($\Delta\mathrm{T}_0$) in the previous lemma is quasilinear by Lemma 2.6, since (ΔT) is. The next lemma gives compatible refinements of (T) when $\mathfrak{e} \prec f - \hat{f}$ in the same way as [ADH 2017, Lemma 14.4.9].

Lemma 6.8. *Suppose that* (T) *is a Tschirnhaus refinement of* (Ê) *and $\mathfrak{e} \prec f - \hat{f}$. Then, with $\hat{g} := \hat{f} - f$ and $\mathfrak{m} := \mathfrak{e}$, the refinement* (TC) *of* (T) *is compatible with* (T).

Proof. Since $\mathfrak{e}$ is the largest algebraic starting monomial for (Ê′),

$$\operatorname{ddeg}_{\preccurlyeq \mathfrak{e}} P_{+f+\hat{g}} = \operatorname{ddeg}_{\preccurlyeq \mathfrak{e}} P_{+\hat{f}} = \operatorname{ddeg} P_{+\hat{f}, \times \mathfrak{e}} = d,$$

and so (TC) has dominant degree d.

As (T) is a Tschirnhaus refinement of (Ê), let $f_0 \in \hat{K}$ be a solution of (6-3) that best approximates $\hat{f}$ and satisfies $f - \hat{f} \sim f_0 - \hat{f}$. Suppose towards a contradiction that $\hat{g}$ is an approximate solution of (ΔT), so by Lemma 6.7, $\hat{g}$ is also an approximate solution of ($\Delta\mathrm{T}_0$). Then by Lemma 5.14, ($\Delta\mathrm{T}_0$) has a solution $y \sim \hat{g} \sim \hat{f} - f_0$. Thus $\Delta P(f_0 + y) = 0$, so $f_0 + y$ is a solution of (6-3), since $f_0 + y \preccurlyeq \mathfrak{f}$. But also

$$f_0 + y - \hat{f} = y - (\hat{f} - f_0) \prec \hat{f} - f_0,$$

contradicting that f_0 best approximates $\hat{f}$. Hence $\hat{g}$ is not an approximate solution of (ΔT), and so (TC) is compatible with (T). □

In fact, the proof above shows that (ΔT_0) has no approximate solution h with $h \sim \hat{f} - f_0$. We now consider the effect of multiplicative conjugation by $\mathfrak{f}$ on the asymptotic differential equations considered so far, as in [ADH 2017, Remark 14.4.10].

Lemma 6.9. *Consider the asymptotic differential equation*

$$(\mathfrak{f}^{-1}\mathrm{E}) \qquad P_{\times\mathfrak{f}}(Y) = 0, \qquad Y \in \mathfrak{f}^{-1}\mathcal{E}$$

over K. Then $(\mathfrak{f}^{-1}\hat{f}, \mathfrak{f}^{-1}\hat{\mathcal{E}}')$ is an unraveller for

$$(\mathfrak{f}^{-1}\hat{\mathrm{E}}) \qquad P_{\times\mathfrak{f}}(Y) = 0, \qquad Y \in \mathfrak{f}^{-1}\hat{\mathcal{E}}$$

over $\hat{K}$, and $\operatorname{ddeg}_{\prec 1}(P_{\times\mathfrak{f}})_{+\mathfrak{f}^{-1}\hat{f}} = d = \operatorname{ddeg}_{\mathfrak{f}^{-1}\hat{\mathcal{E}}} P_{\times\mathfrak{f}}$. *Moreover, if* (T) *is a Tschirnhaus refinement of* (Ê), *then*

$$(\mathfrak{f}^{-1}\mathrm{T}) \qquad (P_{\times\mathfrak{f}})_{+\mathfrak{f}^{-1}f}(Y) = 0, \qquad Y \prec 1$$

is a Tschirnhaus refinement of $(\mathfrak{f}^{-1}\hat{\mathrm{E}})$. *If* (TC) *is a compatible refinement of* (T), *then*

$$(\mathfrak{f}^{-1}\mathrm{TC}) \qquad (P_{\times\mathfrak{f}})_{+\mathfrak{f}^{-1}(f+\hat{g})}(Y) = 0, \qquad Y \preccurlyeq \mathfrak{f}^{-1}\mathfrak{m}$$

is a compatible refinement of $(\mathfrak{f}^{-1}\mathrm{T})$.

Proof. The claims in the second sentence follow directly from Lemma 5.4. The other claims are direct but tedious calculations; however, it is important to recall that $\Delta = (\partial^{\boldsymbol{i}})_{\times\mathfrak{f}}$, so Δ depends on $\mathfrak{f}$, and by [ADH 2017, Lemma 12.8.8],

$$((\partial^{\boldsymbol{i}})_{\times\mathfrak{f}} P)_{\times\mathfrak{f}} = \partial^{\boldsymbol{i}} P_{\times\mathfrak{f}}. \qquad \square$$

6.C. *The slowdown lemma.* In this subsection, we assume that (T) is a Tschirnhaus refinement of (Ê) and (TC) is a compatible refinement of (T). Set $\mathfrak{g} := \mathfrak{d}_{\hat{g}}$, with $\hat{g}$ as in (TC). The main result of this subsection is Lemma 6.11, called the slowdown lemma. A consequence of this, Lemma 6.13, gives the reduction to the special case of Proposition 6.1 considered in Section 6.A. We first prove the following preliminary lemma, which is based on [ADH 2017, Lemma 14.4.12] but has a different proof. Two main differences between these settings play a role here: the change from newton polynomials to dominant parts, and the difference between the choices of Δ. Recall from Section 2.E the coarsening $\preccurlyeq_\phi$ of $\preccurlyeq$ for $\phi \in K^\times$ with $\phi \not\preccurlyeq 1$.

Lemma 6.10. *Suppose that* $\mathfrak{f} = 1$. *Then*

$$\Delta P_{+f}(\hat{g}) \asymp_{\mathfrak{g}} \mathfrak{g}\, \Delta P_{+f}.$$

Proof. Let $f_0 \in \hat{K}$ be a solution of (6-3) that best approximates $\hat{f}$ and satisfies $f - \hat{f} \sim f_0 - \hat{f}$; in particular, $f_0 \sim f \sim \hat{f} \asymp 1$. For this proof, set $Q := \Delta P$.

Since (TC) is compatible with (T), $\hat{g}$ is not an approximate solution of (ΔT), and thus, with $u := \hat{g}/\mathfrak{g}$,

$$D_{Q_{+f,\times\mathfrak{g}}}(\bar{u}) \neq 0.$$

This yields

$$Q_{+f}(\hat{g}) = Q_{+f,\times\mathfrak{g}}(u) \asymp Q_{+f,\times\mathfrak{g}}.$$

Now, since $f - f_0 \prec \mathfrak{g}$, Lemma 2.18(i) gives

$$Q_{+f,\times\mathfrak{g}} = Q_{\times\mathfrak{g},+f/\mathfrak{g}} \sim Q_{\times\mathfrak{g},+f_0/\mathfrak{g}} = Q_{+f_0,\times\mathfrak{g}}.$$

As f_0 is a solution of (6-3), we have

$$\operatorname{mul} Q_{+f_0} = \operatorname{ddeg} Q_{+f_0} = 1$$

by Lemma 6.4. Using Lemma 2.23 and Lemma 2.18(i) again, we get

$$Q_{+f_0,\times\mathfrak{g}} \asymp_{\mathfrak{g}} \mathfrak{g}\, Q_{+f_0} \sim \mathfrak{g}\, Q_{+f}.$$

Finally, we obtain the desired result by combining these steps:

$$Q_{+f}(\hat{g}) \asymp_{\mathfrak{g}} \mathfrak{g}\, Q_{+f}. \qquad \square$$

Using this result, we now turn to the proof of the slowdown lemma, based on [ADH 2017, Lemma 14.4.11]. In its statement and proof, the map $vg \mapsto [vg]$ replaces $vg \mapsto v(g'/g)$ as in Lemma 5.10; this change consequently appears also in Corollary 6.12. The idea, as Aschenbrenner, van den Dries, and van der Hoeven note, is that "the step from (E) to (T) is much larger than the step from (T) to (TC)" [ADH 2017, p. 661 or `arXiv` p. 565].

Lemma 6.11 (slowdown lemma). *With* $\mathfrak{m}$ *the monomial appearing in* (TC), *we have*

$$\left[v\left(\frac{\mathfrak{m}}{\mathfrak{g}}\right)\right] < \left[v\left(\frac{\mathfrak{g}}{\mathfrak{f}}\right)\right].$$

Proof. By Lemma 6.9, we may assume that $\mathfrak{f} = 1$, so $\mathfrak{m} \prec f - \hat{f} \preccurlyeq \mathfrak{g} \prec 1$ and $\Delta = \partial^{\boldsymbol{i}}$. Set $F := P_{+f}$ and note that $\operatorname{ddeg} F_{+\hat{g}} = \operatorname{ddeg} F = d$ by Lemma 2.2(i).

Claim 6.11.1. $$\mathfrak{g}(F_{+\hat{g}})_d \preccurlyeq_{\mathfrak{g}} (F_{+\hat{g}})_{d-1}.$$

Proof of Claim 6.11.1. By Lemma 6.10, we have $\mathfrak{g}\,\partial^{\boldsymbol{i}} F \asymp_{\mathfrak{g}} \partial^{\boldsymbol{i}} F(\hat{g})$, and hence it suffices to show that $(F_{+\hat{g}})_d \asymp \partial^{\boldsymbol{i}} F$ and $\partial^{\boldsymbol{i}} F(\hat{g}) \preccurlyeq (F_{+\hat{g}})_{d-1}$.

By the choice of $\boldsymbol{i}$, we have $\partial^{\boldsymbol{i}} P \asymp P$, so $\partial^{\boldsymbol{i}} F \asymp F_{+\hat{g}}$ by Lemma 2.18(i). As $\operatorname{ddeg} F_{+\hat{g}} = d$, we have $F_{+\hat{g}} \asymp (F_{+\hat{g}})_d$, and thus $(F_{+\hat{g}})_d \asymp \partial^{\boldsymbol{i}} F$. By Taylor expansion, $\partial^{\boldsymbol{i}} F(\hat{g})$ is, up to a factor from $\mathbb{Q}^\times$, the coefficient of $Y^{\boldsymbol{i}}$ in $F_{+\hat{g}}$. Since $|\boldsymbol{i}| = d - 1$, this yields $\partial^{\boldsymbol{i}} F(\hat{g}) \preccurlyeq (F_{+\hat{g}})_{d-1}$. ■

Claim 6.11.2. $$\mathfrak{n} \prec_{\mathfrak{n}} \mathfrak{g} \implies \operatorname{ddeg} F_{+\hat{g},\times\mathfrak{n}} \leqslant d - 1.$$

Proof of Claim 6.11.2. Suppose that $\mathfrak{n} \prec_{\mathfrak{n}} \mathfrak{g}$. Then $\mathfrak{n} \prec 1$, so by Corollary 2.4,

$$\operatorname{ddeg} F_{+\hat{g},\times\mathfrak{n}} \leqslant \operatorname{ddeg} F_{+\hat{g}} = d,$$

and hence it suffices to show that $(F_{+\hat{g},\times\mathfrak{n}})_d \prec_{\mathfrak{n}} (F_{+\hat{g},\times\mathfrak{n}})_{d-1}$. By Lemma 2.22, for all i,

$$(F_{+\hat{g},\times\mathfrak{n}})_i = ((F_{+\hat{g}})_i)_{\times\mathfrak{n}} \asymp_{\mathfrak{n}} \mathfrak{n}^i (F_{+\hat{g}})_i,$$

so it suffices to show that $\mathfrak{n}(F_{+\hat{g}})_d \prec_{\mathfrak{n}} (F_{+\hat{g}})_{d-1}$. First, since $(F_{+\hat{g}})_d \neq 0$, we have $\mathfrak{n}(F_{+\hat{g}})_d \prec_{\mathfrak{n}} \mathfrak{g}(F_{+\hat{g}})_d$. Second, $\mathfrak{n} \prec \mathfrak{g} \prec 1$ implies $[v\mathfrak{g}] \leqslant [v\mathfrak{n}]$, so the first claim and Lemma 2.21 yield $\mathfrak{g}(F_{+\hat{g}})_d \preccurlyeq_{\mathfrak{n}} (F_{+\hat{g}})_{d-1}$. Combining these two relations, we obtain $\mathfrak{n}(F_{+\hat{g}})_d \prec_{\mathfrak{n}} (F_{+\hat{g}})_{d-1}$, as desired. ■

To finish the proof of the lemma, note that $\operatorname{ddeg} F_{+\hat{g},\times\mathfrak{m}} = d$, because (TC) is compatible. Then the second claim gives $\mathfrak{g} \preccurlyeq_{\mathfrak{m}} \mathfrak{m}$, and so $\mathfrak{g} \preccurlyeq_{\mathfrak{g}} \mathfrak{m}$ by Lemma 2.20. But since $\mathfrak{m} \prec \mathfrak{g}$, we must have $\mathfrak{m} \asymp_{\mathfrak{g}} \mathfrak{g}$, giving $[v\mathfrak{m} - v\mathfrak{g}] < [v\mathfrak{g}]$, as desired. □

6.D. *Consequences of the slowdown lemma.* This first consequence, corresponding to [ADH 2017, Corollary 14.4.13], follows immediately from Lemmas 6.8 and 6.11.

Corollary 6.12. *If* (T) *is a Tschirnhaus refinement of* (Ê), *then*

$$\mathfrak{e} \prec \hat{f} - f \implies \left[v\left(\frac{\mathfrak{e}}{\hat{f} - f} \right) \right] < \left[v\left(\frac{\hat{f} - f}{\hat{f}} \right) \right].$$

This next consequence, corresponding to [ADH 2017, Lemma 14.4.14], provides the reduction from Proposition 6.1 to Lemma 6.3.

Lemma 6.13. *Suppose that* (T) *is a Tschirnhaus refinement of* (Ê) *and* $\mathfrak{e} \prec \hat{f} - f$. *Let* $F := P_{+f}$, $\hat{g} := \hat{f} - f$, *and* $\mathfrak{g} := \mathfrak{d}_{\hat{g}}$. *Then the asymptotic differential equation*

$$(\hat{\mathrm{E}}_{\mathfrak{g},\leqslant d}) \qquad F_{\leqslant d}(Y) = 0, \qquad Y \preccurlyeq \mathfrak{g}$$

has dominant degree d. *Moreover, with* $\hat{\mathcal{E}}'_{\mathfrak{g}} := \{y \in \hat{\mathcal{E}}' : y \prec \mathfrak{g}\}$, $(\hat{g}, \hat{\mathcal{E}}'_{\mathfrak{g}})$ *is an unraveller for* $(\hat{\mathrm{E}}_{\mathfrak{g},\leqslant d})$ *and* $\mathfrak{e}$ *is the largest algebraic starting monomial for the unravelled asymptotic differential equation*

$$(\hat{\mathrm{E}}'_{\mathfrak{g},\leqslant d}) \qquad (F_{\leqslant d})_{+\hat{g}}(Y) = 0, \qquad Y \in \hat{\mathcal{E}}'_{\mathfrak{g}}$$

over $\hat{K}$.

Proof. This follows from Corollary 6.12 by applying Lemma 5.10 with $\hat{K}$, $\hat{f}$, f, $\hat{g}$, $\hat{\mathcal{E}}$, $\hat{\mathcal{E}}'$, and $\hat{\mathcal{E}}'_{\mathfrak{g}}$ in the roles of K, f, $f - g$, g, $\mathcal{E}$, $\mathcal{E}'$, and $\mathcal{E}'_g$, respectively. □

6.E. ***Proposition 6.1 and its consequence.*** Finally, we return to the proof of the main proposition of this section. Recall the following statement:

Proposition 6.1. *There exists $f \in \hat{K}$ such that one of the following holds*:

(i) $\hat{f} - f \preccurlyeq \mathfrak{e}$ *and* $A(f) = 0$ *for some* $A \in K\{Y\}$ *with* $c(A) < c(P)$ *and* $\deg A = 1$;

(ii) $\hat{f} \sim f$, $\hat{f} - a \preccurlyeq f - a$ *for all* $a \in K$, *and* $A(f) = 0$ *for some* $A \in K\{Y\}$ *with* $c(A) < c(P)$ *and* $\operatorname{ddeg} A_{\times f} = 1$.

Proof. As noted already, if $\mathfrak{e} \succcurlyeq \mathfrak{f}$, then case (i) holds with $f := 0$ and $A := Y$, so suppose that $\mathfrak{e} \prec \mathfrak{f}$. By Lemma 6.5, $\hat{f}$ is an approximate solution of (6-3), so by Lemmas 5.14 and 5.13, we have a solution $f_0 \sim \hat{f}$ in $\hat{K}$ of (6-3) that best approximates $\hat{f}$. If $\hat{f} - a \preccurlyeq f_0 - a$ for all $a \in K$, then case (ii) holds with $f := f_0$ and $A := \Delta P$. Now suppose to the contrary that we have $f \in K$ with $\hat{f} - f \succ f_0 - f$. That is, $f_0 - \hat{f} \sim f - \hat{f}$, so in view of $f_0 \sim \hat{f}$, we have $f \sim \hat{f}$. Hence (T) is a Tschirnhaus refinement of (Ê). We are going to show that then case (i) holds.

If $\hat{f} - f \preccurlyeq \mathfrak{e}$, then case (i) holds with $A := Y - f$, so for the rest of the proof, assume that $\mathfrak{e} \prec \hat{f} - f$, and set $F := P_{+f}$, $\hat{g} := \hat{f} - f$, and $\mathfrak{g} := \mathfrak{d}_{\hat{g}}$. This puts us in the situation of the previous lemma, so $(\hat{\mathrm{E}}_{\mathfrak{g},\leqslant d})$ has dominant degree d and $(\hat{g}, \hat{\mathcal{E}}'_{\mathfrak{g}})$ is an unraveller for $(\hat{\mathrm{E}}_{\mathfrak{g},\leqslant d})$. In particular, $\operatorname{ddeg}_{\prec \hat{g}}(F_{\leqslant d})_{+\hat{g}} = d$, since

$$\operatorname{ddeg}_{\prec \hat{g}}(F_{\leqslant d})_{+\hat{g}} \geqslant \operatorname{ddeg}_{\hat{\mathcal{E}}'_{\mathfrak{g}}}(F_{\leqslant d})_{+\hat{g}} = d.$$

Also, $\mathfrak{e}$ is the largest algebraic starting monomial for $(\hat{\mathrm{E}}'_{\mathfrak{g},\leqslant d})$. Now since $f \in K$, we can view $(\hat{\mathrm{E}}_{\mathfrak{g},\leqslant d})$ as an asymptotic differential equation over K. We also have $\deg F_{\leqslant d} = d$ and $\operatorname{mul}(F_{\leqslant d})_{+\hat{g}} < d$, since otherwise $(F_{\leqslant d})_{+\hat{g}}$ would be homogeneous and so not have any algebraic starting monomials. Thus with $(\hat{\mathrm{E}}_{\mathfrak{g},\leqslant d})$ in place of (E) and $(\hat{g}, \hat{\mathcal{E}}'_{\mathfrak{g}})$ in place of $(\hat{f}, \hat{\mathcal{E}}')$, Lemma 6.3 applies. Hence we have $g \in \hat{K}$ and $B \in K\{Y\}$ such that $\hat{g} - g \preccurlyeq \mathfrak{e}$, $B(g) = 0$, $c(B) < c(F_{\leqslant d})$, and $\deg B = 1$. Finally, case (i) holds with $f + g$ in place of f and with $A := B_{-f}$, completing the proof. □

In fact, if K is r-d-henselian, then the $f \in \hat{K}$ in Proposition 6.1 actually lies in K. This follows easily from [ADH 2017, Proposition 7.5.6], just as [ADH 2017, Corollary 14.4.16] follows from [ADH 2017, Lemma 14.1.8]. We do not use Proposition 6.1 directly in the proof of Proposition 3.1, but rather this corollary concerning pc-sequences, corresponding to [ADH 2017, Corollary 14.4.15]:

Corollary 6.14. *Suppose that (a_ρ) is a divergent pc-sequence in K with pseudolimit $\hat{f} \in \hat{K}$ and minimal d-polynomial P over K. Then there exist $f \in \hat{K}$ and $A \in K\{Y\}$ such that $\hat{f} - f \preccurlyeq \mathfrak{e}$, $A(f) = 0$, $c(A) < c(P)$, and* $\deg A = 1$.

Proof. Suppose towards a contradiction that there are no such f and A. Then Proposition 6.1 gives instead $f \in \hat{K}$ and $A \in K\{Y\}^{\neq}$ such that $\hat{f} - a \preccurlyeq f - a$ for all $a \in K$, $A(f) = 0$, and $c(A) < c(P)$. Since (a_ρ) has no pseudolimit in K, $\hat{f} \notin K$, and so $f \notin K$. Hence we may take a divergent pc-sequence (b_σ) in K such that $b_\sigma \rightsquigarrow f$. Since $\hat{f} - b_\sigma \preccurlyeq f - b_\sigma$ for all σ, we have $b_\sigma \rightsquigarrow \hat{f}$. The pc-sequences (a_ρ) and (b_σ) have no pseudolimit in K but the common pseudolimit $\hat{f} \in \hat{K}$, and hence they are equivalent by [ADH 2017, Corollary 2.2.20]. Thus $a_\rho \rightsquigarrow f$, so applying Lemma 2.11 to A and f contradicts the minimality of P. □

7. Proof of Proposition 3.1

In this section, we prove the main proposition, derived from the work of the previous sections, thus completing the proof of the main results. Its proof is based on that of [ADH 2017, Proposition 14.5.1].

Proposition 3.1. *Suppose that K is asymptotic, Γ is divisible, and $\boldsymbol{k}$ is r-linearly surjective. Let (a_ρ) be a pc-sequence in K with minimal d-polynomial G over K of order at most r. Then* $\operatorname{ddeg}_{\boldsymbol{a}} G = 1$.

Proof. Let $d := \operatorname{ddeg}_{\boldsymbol{a}} G$. We may assume that (a_ρ) has no pseudolimit in K, as otherwise, up to scaling, G is of the form $Y - a$ for some pseudolimit a of (a_ρ), and hence $d = 1$. We may also assume that $r \geqslant 1$, since the case $r = 0$ is handled by the analogous fact for valued fields of equicharacteristic 0 (see [ADH 2017, Proposition 3.3.19]). By Zorn's lemma, we may take a d-algebraically maximal immediate extension $\hat{K}$ of K. By the proof of [ADH 2017, Theorem 7.0.1], $\hat{K}$ is r-d-henselian. Note that as an immediate extension of K, $\hat{K}$ is also asymptotic by [ADH 2017, Lemmas 9.4.2 and 9.4.5].

Now, take $\ell \in \hat{K}$ such that $a_\rho \rightsquigarrow \ell$, so G is an element of minimal complexity of $Z(K, \ell)$ by Corollary 2.17. Lemma 2.16 gives $d \geqslant 1$, as well as $a \in K$ and $\mathfrak{v} \in K^\times$ such that $a - \ell \prec \mathfrak{v}$ and $\operatorname{ddeg}_{\prec\mathfrak{v}} G_{+a} = d$. Towards a contradiction, suppose that $d \geqslant 2$. Lemma 5.8 then yields an unraveller $(\hat{f}, \hat{\mathcal{E}})$ for the asymptotic differential equation

$$G_{+a}(Y) = 0, \qquad Y \prec \mathfrak{v} \tag{7-1}$$

over $\hat{K}$ such that

(i) $\hat{f} \neq 0$,

(ii) $\operatorname{ddeg}_{\prec \hat{f}} G_{+a+\hat{f}} = d$,

(iii) $a_\rho \rightsquigarrow a + \hat{f} + g$ for all $g \in \hat{\mathcal{E}} \cup \{0\}$,

(iv) $\operatorname{mul} G_{+a+\hat{f}} < d$,

where (iv) follows from (iii) by Lemma 5.7.

Suppose first that K has a monomial group. Consider the pc-sequence $(a_\rho - a)$ with minimal d-polynomial $P := G_{+a}$ over K. Since $(\hat{f}, \hat{\mathcal{E}})$ is an unraveller for (7-1), $\operatorname{ddeg}_{\hat{\mathcal{E}}} P_{+\hat{f}} = d > \operatorname{mul} P_{+\hat{f}}$ by (iv), so let $\mathfrak{e}$ be the largest algebraic starting monomial for the asymptotic differential equation

$$P_{+\hat{f}}(Y) = 0, \qquad Y \in \hat{\mathcal{E}} \tag{7-2}$$

over $\hat{K}$ by Proposition 4.4. Hence all the assumptions of the previous section are satisfied (with (7-1) and (7-2) in the roles of (Ê) and (Ê′), respectively), so applying Corollary 6.14 to $(a_\rho - a)$ and P yields $f \in \hat{K}$ and $A \in K\{Y\}^{\neq}$ such that $\hat{f} - f \preccurlyeq \mathfrak{e}$, $A(f) = 0$, and $c(A) < c(P)$. Since $\mathfrak{e}$ is an algebraic starting monomial for (7-2), we have $\mathfrak{e} \in \hat{\mathcal{E}}$, and so $f - \hat{f} \in \hat{\mathcal{E}} \cup \{0\}$. But then $a_\rho - a \rightsquigarrow f$ by (iii), so applying Lemma 2.11 to A and f contradicts the minimality of P.

Finally, we reduce to the case that K has a monomial group. Consider $\hat{K}$ as a valued differential field with a predicate for K and pass to an $\aleph_1$-saturated elementary extension of this structure. In particular, the new K has a monomial group [ADH 2017, Lemma 3.3.39]. In doing this, we preserve all the relevant first order properties: small derivation, r-linearly surjective differential residue field, divisible value group, asymptoticity, r-d-henselianity of $\hat{K}$, and that $G \in Z(K, \ell)$ but $H \notin Z(K, \ell)$ for all $H \in K\{Y\}$ with $c(H) < c(G)$.

However, it is possible that $\hat{K}$ is no longer d-algebraically maximal, in which case we pass to a d-algebraically maximal immediate extension of $\hat{K}$ (and hence of K). It is also possible that (a_ρ) is no longer divergent in K, in which case we replace (a_ρ) with a divergent pc-sequence (b_σ) in K with $b_\sigma \rightsquigarrow \ell$. By Corollary 2.17, G is a minimal d-polynomial of (b_σ) over K, and by Lemma 2.16, $\operatorname{ddeg}_{\boldsymbol{b}} G = d$, where $\boldsymbol{b} := c_K(b_\sigma)$. By the argument above used in this new structure, $d = 1$, as desired. □

Acknowledgements

This research was supported by an NSERC Postgraduate Scholarship. Thanks are due to Lou van den Dries for many helpful discussions and for comments on a draft of this paper, as well as to the reviewer for their careful reading of the manuscript and numerous comments and suggestions that have improved the clarity and readability of the paper.

References

[ADH 2017] M. Aschenbrenner, L. van den Dries, and J. van der Hoeven, *Asymptotic differential algebra and model theory of transseries*, Ann. of Math. Stud. **195**, Princeton Univ. Press, 2017. MR Zbl

[ADH 2018] M. Aschenbrenner, L. van den Dries, and J. van der Hoeven, "Maximal immediate extensions of valued differential fields", *Proc. Lond. Math. Soc.* (3) **117**:2 (2018), 376–406. MR Zbl

[van den Dries and Pynn-Coates 2019] L. van den Dries and N. Pynn-Coates, "On the uniqueness of maximal immediate extensions of valued differential fields", *J. Algebra* **519** (2019), 87–100. MR Zbl

[Kaplansky 1942] I. Kaplansky, "Maximal fields with valuations", *Duke Math. J.* **9** (1942), 303–321. MR Zbl

[Kuhlmann 2011] F.-V. Kuhlmann, "Maps on ultrametric spaces, Hensel's lemma, and differential equations over valued fields", *Comm. Algebra* **39**:5 (2011), 1730–1776. MR Zbl

[Matusinski 2014] M. Matusinski, "On generalized series fields and exponential-logarithmic series fields with derivations", pp. 350–372 in *Valuation theory in interaction* (Segovia/El Escorial, Spain, 2011), edited by A. Campillo et al., Eur. Math. Soc., Zürich, 2014. MR Zbl

[Pynn-Coates 2019] N. Pynn-Coates, "Newtonian valued differential fields with arbitrary value group", *Comm. Algebra* **47**:7 (2019), 2766–2776. MR Zbl

[Rosenlicht 1980] M. Rosenlicht, "Differential valuations", *Pacific J. Math.* **86**:1 (1980), 301–319. MR Zbl

[Scanlon 2000] T. Scanlon, "A model complete theory of valued D-fields", *J. Symbolic Logic* **65**:4 (2000), 1758–1784. MR Zbl

Received August 7, 2018. Revised March 25, 2020.

NIGEL PYNN-COATES
DEPARTMENT OF MATHEMATICS
THE OHIO STATE UNIVERSITY
COLUMBUS, OH
UNITED STATES

pynn-coates.1@osu.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 308, No. 1, 2020
https://doi.org/10.2140/pjm.2020.308.207

CONJUGACY CLASSES OF p-ELEMENTS AND NORMAL p-COMPLEMENTS

HUNG P. TONG-VIET

We study the structure of finite groups with a large number of conjugacy classes of p-elements for some prime p. As a consequence, we obtain some new criteria for the existence of normal p-complements in finite groups.

1. Introduction

Let p be a prime. Let G be a finite group and let P be a Sylow p-subgroup of G. Denote by $k(G)$ and $k_p(G)$ the number of conjugacy classes of G and the number of conjugacy classes of p-elements of G, respectively. By Sylow's theorem, we can choose a complete set Γ of representatives for the conjugacy classes of p-elements of G in such a way that $\Gamma \subseteq P$. This yields that $k_p(G) \leq k(P)$. Also $k_p(G) \geq 2$ unless G is a p'-group. Hence if p divides $|G|$, then $2 \leq k_p(G) \leq k(P) \leq |P|$. In [Külshammer et al. 2014], the authors study finite groups G with $k_p(G) = 2$. They show that the Sylow p-subgroup P of such a group G must be either elementary abelian or extra-special of order p^3. In this paper, we will look at the case when $k_p(G)$ is large in comparison to $|P|$.

Recall that a subgroup N of a finite group G is called a normal p-complement of G if N is a normal subgroup of G whose order is relatively prime to p and whose index is a power of p. A finite group G is said to be p-nilpotent if it has a normal p-complement. A classical result in group theory states that a finite group G is p-nilpotent if and only if P controls its own fusion in G. (See [Isaacs 2008, Theorem 5.25] and the definitions in Section 2). The latter condition is equivalent to $x^G \cap P = x^P$ for every $x \in P$, which is equivalent to the condition $k_p(G) = k(P)$. Thus G is p-nilpotent if and only if $k_p(G) = k(P)$. If we assume that $k_p(G) = |P|$, then $k_p(G) = k(P) = |P|$; hence G is p-nilpotent and has an abelian Sylow p-subgroup. So, we may ask whether G is still p-nilpotent, if $k_p(G)/|P|$ is close to 1.

It turns out that the fraction $k_p(G)/|P|$ is related to the commuting probability $d(G)$ of a finite group G, which is defined to be the probability that two randomly chosen elements of G commute. Gustafson [1973] shows that $d(G) = k(G)/|G|$. The invariant $d(G)$ is also called the commutativity degree of G.

MSC2020: primary 20E45; secondary 20D10, 20D20.
Keywords: conjugacy classes, p-elements, normal p-complements.

Here is our first result for the prime $p = 2$.

Theorem A. *Let G be a finite group and let P be a Sylow 2-subgroup of G. Then $k_2(G) > |P|/2$ if and only if G has a normal 2-complement and $k(P) > |P|/2$.*

Clearly, any finite group with a normal 2-complement is solvable by the Feit–Thompson theorem. Also, the Sylow 2-subgroup P in Theorem A is nilpotent of class at most 2. (See Lemma 2.7.) Theorem A does not hold if we allow equality. For example, if $G = \mathbf{A}_4$ and $P \in \mathrm{Syl}_2(\mathrm{G})$, then $k_2(G) = 2$ and $|P| = 4$, so $k_2(G) = |P|/2$ but G is not 2-nilpotent. Also, we cannot replace 2 by an odd prime. Indeed, if $G = \mathbf{A}_5$ and $P \in \mathrm{Syl}_3(\mathrm{G})$, then $k_3(G) = 2$ and $|P| = 3$; hence $k_3(G) = \frac{2}{3}|P| > \frac{1}{2}|P|$ but G is not 3-nilpotent.

In view of Lemma 2.8, to investigate the structure of finite groups G with $k_p(G)/|P|$ a specified constant, we may assume that $\mathbf{O}_{p'}(G) = 1$.

Theorem B. *Let G be a finite group and let $P \in \mathrm{Syl}_2(\mathrm{G})$. Suppose that $\mathbf{O}_{2'}(G) = 1$ and $k_2(G) = |P|/2$. Then*

(1) *$G/\mathbf{Z}(G) \cong \mathbf{A}_4$ or $\mathbf{S}_4$; or*

(2) *$G/\mathbf{Z}(G)$ is an almost simple group with a nonabelian simple socle isomorphic to $\mathrm{PSL}_2(q)$ with $3 < q \equiv 3, 5 \pmod 8$.*

Let π be a set of primes. Let $k_\pi(G)$ be the number of conjugacy classes of π-elements of G. Let $|G|_\pi$ be the π-part of the order $|G|$ of G. Define $d_\pi(G)$ to be $k_\pi(G)/|G|_\pi$. If $\pi = \{p\}$, then we write $d_p(G)$ and $k_p(G)$ instead of $d_{\{p\}}(G)$ or $k_{\{p\}}(G)$. We now investigate the structure of finite groups G with $d_\pi(G) > \frac{1}{2}$, where π is a set of primes containing 2.

Theorem C. *Let G be a finite group and let π be a set of primes with $2 \in \pi$. Let $\sigma = \pi \setminus \{2\}$. Suppose that $d_\pi(G) > \frac{1}{2}$. Then G has a normal π-complement and an abelian Hall σ-subgroup.*

We should point out that our proofs of Theorems A–C do not depend on the classification of finite simple groups.

For odd primes p, we obtain the following result. Here our proof depends on the odd version of Glauberman's Z^*-theorem and thus depends on the classification of finite simple groups.

Theorem D. *Let G be a finite group and let p be an odd prime. Then $d_p(G) > (p+1)/2p$ if and only if G has a normal p-complement and an abelian Sylow p-subgroup.*

This bound cannot be improved since $d_p(\mathrm{D}_{2p}) = (p+1)/2p$ but D_{2p} is not p-nilpotent, where p is an odd prime. For nonsolvable examples, let $f \geq 2$ be an integer and p be a prime such that $4^f - 1$ is divisible by p but not by p^2. Then $d_p(\mathrm{PSL}_2(2^f)) = (p+1)/2p$.

Theorem E. *Let G be a finite group and let π be a set of odd primes. Let p be the smallest prime in π. Suppose that $d_\pi(G) > (p+1)/2p$. Then G has a normal π-complement and an abelian Hall π-subgroup.*

In [Maróti and Nguyen 2014], the authors show that if $d_\pi(G) > \frac{5}{8}$, then $d_\pi(G) = 1$ or $\frac{2}{3}$. They also study the structure of finite groups G such that $\boldsymbol{O}_{3'}(G) = 1$ and $d_3(G) = \frac{2}{3}$. Thus if $p = 3$ in Theorem D, then our results follow immediately from their results. However, if $p \geq 5$, then $(p+1)/2p < \frac{5}{8}$. Hence our Theorems C and E above improve their Theorem 1 and finally our last result includes Theorem 2 in [Maróti and Nguyen 2014].

Theorem F. *Let G be a finite group and let p be an odd prime. Let P be a Sylow p-subgroup of G. Suppose that $\boldsymbol{O}_{p'}(G) = 1$ and $d_p(G) = (p+1)/2p$. Then P is abelian, $\boldsymbol{N}_G(P)/\boldsymbol{C}_G(P)$ has order 2, $[P, \boldsymbol{N}_G(P)]$ has order p and $G \cong A \times B$, where B is an abelian p-group and A is either a dihedral group of order $2p$ or an almost simple group with a Sylow p-subgroup of order p contained in the socle of A.*

The paper is organized as follows. We collect some results needed for the proofs of the main theorems in Section 2. We prove Theorems A–C in Section 3 and prove Theorems D–F in Section 4.

2. Control of fusion and Glauberman's Z^*-theorem

Let G be a finite group and let $K \leq H \leq G$ be subgroups of G. We say that H *controls G-fusion in K* if and only if every pair of G-conjugate elements of K are H-conjugate; that is, if $x, x^g \in K$ for some $g \in G$, then $x^g = x^h$ for some $h \in H$. Let p be a prime and let H be a subgroup of G. We say that H *controls p-fusion* in G if H contains a Sylow p-subgroup P of G and H controls G-fusion in P. We first recall some classical results on the existence of normal p-complements as well as the control of fusion in finite groups.

Lemma 2.1. *Let G be a finite group and let P be a Sylow p-subgroup of G for some prime p.*

(1) *$\boldsymbol{N}_G(P)$ controls G-fusion in $\boldsymbol{C}_G(P)$.*

(2) *If $P \subseteq \boldsymbol{Z}(\boldsymbol{N}_G(P))$, then G has a normal p-complement.*

(3) *G has a normal p-complement if and only if P controls its own fusion in G.*

Proof. These are well known results; for proofs, see Lemma 5.12 and Theorems 5.13 and 5.25 in [Isaacs 2008]. □

Parts (1) and (2) above are known as Burnside's lemma and Burnside's normal p-complement theorem, respectively. Here are some obvious consequences of the lemma.

Corollary 2.2. *Let G be a finite group and let P be a Sylow p-subgroup of G for some prime p.*

(1) $k_p(G) \le k(P)$, *and equality holds if and only if G has a normal p-complement.*

(2) $k_p(G) = |P|$, *or equivalently* $d_p(G) = 1$, *if and only if G has a normal p-complement and an abelian Sylow p-subgroup.*

Note that part (1) of the corollary is equivalent to the statement that $d_p(G) \le d(P)$ and equality holds if and only if G has a normal p-complement. The following result is a consequence of the definitions above and Sylow's theorem.

Lemma 2.3. *Let G be a finite group and let P be a Sylow p-subgroup of G for some prime p. Let $x \in P$. Then $x^G \cap P = \{x\}$ if and only if $\mathbf{C}_G(x)$ controls p-fusion in G.*

Proof. Let $x \in P$. Assume that $x^G \cap P = \{x\}$. We claim that $\mathbf{C}_G(x)$ controls p-fusion in G. Since $x^P \subseteq x^G \cap P = \{x\}$, we see that $x \in \mathbf{Z}(P)$ and thus $P \le \mathbf{C}_G(x)$. Now assume that $y, y^g \in P$ for some $g \in G$. We need to show that $y^g = y^h$ for some $h \in \mathbf{C}_G(x)$. We have that $\{y, y^g\} \subseteq P \subseteq \mathbf{C}_G(x)$, which implies that $\{x, x^{g^{-1}}\} \subseteq \mathbf{C}_G(y)$. Let U be a Sylow p-subgroup of $\mathbf{C}_G(y)$ containing x. By Sylow's theorem, $U \le P^t$ for some $t \in G$. It follows that $x^{t^{-1}} \in P \cap x^G = \{x\}$; hence $x^{t^{-1}} = x$, so $t \in \mathbf{C}_G(x)$. Now $x^{g^{-1}} \in U^c$ for some $c \in \mathbf{C}_G(y)$ as $x^{g^{-1}} \in \mathbf{C}_G(y)$ is a p-element. We now have that $x^{g^{-1}c^{-1}t^{-1}} \in P$ and thus $x^{g^{-1}c^{-1}t^{-1}} = x$, which implies that $g^{-1}c^{-1} \in \mathbf{C}_G(x)$. Therefore $cg = h \in \mathbf{C}_G(x)$. Now $y^g = y^{cg} = y^h$ as wanted.

For the converse, let $P_1 \in \mathrm{Syl_p(G)}$ and assume that $P_1 \subseteq \mathbf{C}_G(x)$ and that $\mathbf{C}_G(x)$ controls G-fusion in P_1. It follows that $x \in P_1$. By Sylow's theorem, $P = P_1^t$ for some $t \in G$. Since $x \in P$, $x^{t^{-1}} \in P_1 \le \mathbf{C}_G(x)$. As $\mathbf{C}_G(x)$ controls G-fusion in P_1, it follows that $x^{t^{-1}} = x^h$ for some $h \in \mathbf{C}_G(x)$. Hence $x^{t^{-1}} = x^h = x$ and so $t \in \mathbf{C}_G(x)$. In particular, $P = P_1^t \subseteq \mathbf{C}_G(x)$. Finally, if $x^g \in P$ for some $g \in G$, then $x^g = x^h$ for some $h \in \mathbf{C}_G(x)$ and so $x^g = x^h = x$. Therefore $x^G \cap P = \{x\}$. ☐

For a finite group G and a prime p, we define $\mathbf{Z}_p^*(G)$ to be the normal subgroup of G such that $\mathbf{Z}_p^*(G)/\mathbf{O}_{p'}(G) = \mathbf{Z}(G/\mathbf{O}_{p'}(G))$.

We first state the original form of Glauberman's Z^*-theorem, whose proof does not depend on the classification of finite simple groups.

Lemma 2.4 (Glauberman's Z^*-theorem). *Let G be a finite group and let P be a Sylow 2-subgroup of G. If $x \in P$ and $x^G \cap P = \{x\}$, then $x \in \mathbf{Z}_2^*(G)$.*

Proof. This is a restatement of Theorem 3 in [Glauberman 1966]. ☐

The odd version of Glauberman's Z^*-theorem, which is called Glauberman's Z_p^*-theorem, says that if $x \in P$ is an element of order p and $x^G \cap P = \{x\}$, then $x \in \mathbf{Z}_p^*(G)$. The proof of this theorem depends on the classification (for a sketch proof, see [Guralnick and Robinson 1993, Theorem 4.1]). By Sylow's theorem, it is easy to see that if $x^G \cap P = \{x\}$ then x does not commute with any G-conjugate $x^g \neq x$ of x. Finally, we note that the conclusion of Glauberman's Z_p^*-theorem can be written as $G = \boldsymbol{C}_G(x)\boldsymbol{O}_{p'}(G)$.

To use Glauberman's Z_p^*-theorem for an arbitrary p-element $x \in P$ which is not of prime order satisfying $x^G \cap P = \{x\}$, we need the following lemma.

Lemma 2.5. *Let G be a finite group and let P be a Sylow p-subgroup of G for some prime p. Let $x \in P$. If $x^G \cap P = \{x\}$, then $y^G \cap P = \{y\}$ for every $y \in \langle x \rangle$.*

Proof. Suppose that $x^G \cap P = \{x\}$ and $y \in \langle x \rangle$. Then $P \leq \boldsymbol{C}_G(x) \leq \boldsymbol{C}_G(y)$. By Lemma 2.3, we need to show that $\boldsymbol{C}_G(y)$ controls G-fusion in P. Let $z, z^g \in P$ for some $g \in G$. By Lemma 2.3, $\boldsymbol{C}_G(x)$ controls p-fusion in G so $z^g = z^t$ for some $t \in \boldsymbol{C}_G(x)$. As $\boldsymbol{C}_G(x) \leq \boldsymbol{C}_G(y)$, we have $t \in \boldsymbol{C}_G(y)$ and the claim follows. □

We will need the following results.

Lemma 2.6. *Let G be a finite group and let π be a nonempty set of primes.*

(1) *If $\mu \subseteq \pi$ is a nonempty subset, then $d_\pi(G) \leq d_\mu(G) \leq 1$.*

(2) *If $N \trianglelefteq G$, then $d_\pi(G) \leq d_\pi(G/N)d_\pi(N)$.*

(3) *If G is a nonabelian p-group for some prime p, then $d(G) < (p+1)/p^2$.*

(4) *If G does not have a normal Sylow p-subgroup for some prime p, then $d(G) \leq 1/p$.*

Proof. Part (1) can be found in [Maróti and Nguyen 2014, Proposition 5] and Part (2) is Lemma 2.3 in [Fulman and Guralnick 2012]. Finally, the last two parts can be found in Lemma 2 in [Guralnick and Robinson 2006]. □

Finite groups G with $d(G) \geq \frac{1}{2}$ were classified by Lescot [1995; 2001]. To state the result, we need the following notation. For any integer $m \geq 1$, denote by G_m the group defined by

$$G_m = \langle a, b : a^3 = b^{2^m} = 1, a^b = a^{-1} \rangle.$$

Note that $G_1 \cong S_3$. We have that $|G_m| = 3 \cdot 2^m$, $\boldsymbol{Z}(G_m) = \langle b^2 \rangle$, $G_m' = \langle a \rangle$, and $G_m/\boldsymbol{Z}(G_m) \cong S_3$.

Lemma 2.7. *Let G be a finite group. Then $d(G) \geq \frac{1}{2}$ if and only if one of the following holds*:

(i) *G is abelian and $d(G) = 1$.*

(ii) *$G \cong P \times A$, where A is abelian of odd order and P is a Sylow 2-subgroup of G with $|G'| = |P'| = 2$ and $d(G) = d(P) = (1 + 4^{-m})/2$ and $G/\boldsymbol{Z}(G)$ is elementary abelian of order 4^m for some integer $m \geq 1$. Further, $\frac{1}{2} < d(G) \leq \frac{5}{8}$.*

(iii) *$G \cong G_m \times A$ and $d(G) = \frac{1}{2}$, where A is abelian and $m \geq 1$.*

Proof. This is a combination of Theorem 3.1 in [Lescot 2001] and Corollary 3.2 in [Lescot 1995]. □

It follows from Lemma 2.7 that there is no 2-group G with $d(G) = \frac{1}{2}$. Also, if G is of odd order with $d(G) \geq \frac{1}{2}$, then G is abelian and $d(G) = 1$.

Lemma 2.8. *Let G be a finite group and let p be a prime. If $N \trianglelefteq G$ is a p'-subgroup, then $k_p(G) = k_p(G/N)$ and so $d_p(G) = d_p(G/N)$.*

Proof. Let N be a normal p'-subgroup of G and let $P \in \mathrm{Syl_p(G)}$. Write $\overline{G} = G/N$. Since $p \nmid |N|$, the Sylow p-subgroups of G and $\overline{G}$ have the same order, thus it suffices to show that $k_p(G) = k_p(\overline{G})$. By Lemma 2.6(2), we have $d_p(G) \leq d_p(\overline{G})$ since $d_p(N) \leq 1$. It follows that $k_p(G) \leq k_p(\overline{G})$. The reverse inequality is obvious. □

3. Conjugacy classes of 2-elements

We will prove Theorems A, B and C in this section. Recall that a p-group P is said to be extra-special if $P' = \Phi(P) = \mathbf{Z}(P)$ and $|\mathbf{Z}(P)| = p$. The following lemma is key to our proofs.

Lemma 3.1. *Let G be a finite group and let P be a Sylow 2-subgroup of G. Suppose that $\boldsymbol{O}_{2'}(G) = \mathbf{Z}(G) = 1$ and $|P| > 1$. Then $k_2(G) \leq |P|/2$ and $d_2(G) \leq \frac{1}{2}$. Moreover, if $k_2(G) = |P|/2$, then one of the following holds:*

(1) *$k_2(G) = 2$ and P is elementary abelian of order 4.*

(2) *$k_2(G) > 2$ and P is an extra-special group of order 2^{1+2m} for some integer $m \geq 1$ with $\mathbf{Z}(P) = \langle z \rangle$ a cyclic group of order 2. Moreover, $|z^G \cap P| = 3$ and for any $1 \neq y \in P$ with $y \notin z^G$, $|y^G \cap P| = 2$.*

Proof. The hypothesis of the lemma implies that $\mathbf{Z}_2^*(G) = 1$. Let P be a Sylow 2-subgroup of G. By Lemma 2.4, if $1 \neq x \in P$, then $|x^G \cap P| \geq 2$. Let $k = k_2(G) - 1$ be the number of nontrivial conjugacy classes of 2-elements in G. Clearly, we can choose a complete set $\Gamma = \{x_i\}_{i=1}^k$ of representatives for all nontrivial conjugacy classes of 2-elements in G such that $\Gamma \subseteq P \setminus \{1\}$. Notice that $k \geq 1$ as otherwise P is trivial.

Observe that $|x_i^G \cap P| \geq 2$ for all i with $1 \leq i \leq k$, and $P \setminus \{1\} = \bigcup_{i=1}^k x_i^G \cap P$. We have that

$$|P \setminus \{1\}| = \sum_{i=1}^{k} |x_i^G \cap P| \geq \sum_{i=1}^{k} 2 = 2k.$$

Hence $|P| - 1 \geq 2k$ and so $2k \leq |P| - 2$ since $|P| - 1$ is odd. Therefore $k_2(G) = k + 1 \leq |P|/2$ and $d_2(G) \leq \frac{1}{2}$ as wanted.

Next, assume that $k_2(G) = |P|/2$. We know that $|x_i^G \cap P| \geq 2$ for all $i = 1, 2, \ldots, k$. If $k = 1$, then $k_2(G) = 2$ and $|P| = 4$, so $|x_1^G \cap P| = 3$ and P is elementary abelian of order 4. Thus part (1) holds. Assume that $k \geq 2$. Since $|x_i^G \cap P| \geq 2$

for every i and $k = |P|/2 - 1$, we obtain that $|x_j^G \cap P| = 3$ for a unique index j and $|x_i^G \cap P| = 2$ for all $1 \le i \ne j \le k$. So we may assume that $|x_1^G \cap P| = 3$ and $|x_i^G \cap P| = 2$ for $2 \le i \le k$.

By Corollary 2.2, we have $\frac{1}{2} = d_2(G) \le d(P)$ and thus either P is abelian or $|P'| = 2,\ d(P) = (1 + 4^{-m})/2$ and $P/\mathbf{Z}(P)$ is elementary abelian of order 4^m by Lemma 2.7. We claim that P is nonabelian. By way of contradiction, assume that P is abelian. Let $H = \mathbf{N}_G(P)$. By Lemma 2.1(1), H controls G-fusion in P (since P is abelian) and so $x_k^G \cap P = x_k^H$. Moreover, as P is abelian, $P \le \mathbf{C}_H(x_k) \le H$ and hence $|x_k^H| \ge 1$ is odd. However $|x_k^H| = |x_k^G \cap P| = 2$ by the result in the previous paragraph, which is a contradiction.

Next, we claim that P is extra-special. It suffices to show that $\mathbf{Z}(P) = P'$. Write $P' = \langle z \rangle$. Since $|P'| = 2$, we have $P' \le \mathbf{Z}(H) \cap \mathbf{Z}(P)$. Let $1 \ne u \in \mathbf{Z}(P)$. We claim that $|u^G \cap P| = 3$. Assume by contradiction that $u^G \cap P = \{u, v\}$, where $u \ne v \in P$. If $v \in \mathbf{Z}(P)$, then $v = u^h$ for some $h \in H$ by Lemma 2.1(1). It follows that $|u^H| = 2$, which is impossible as $P \le \mathbf{C}_H(u) \le H$. Thus $v \notin \mathbf{Z}(P)$. Now $u^G \cap P = v^G \cap P = \{u, v\}$. Since $v \in P \setminus \mathbf{Z}(P)$, we have $|v^P| > 1$ whence $v^P = \{u, v\}$. In particular $u = v^t$ for some $t \in P$. Hence $v = u^{t^{-1}} = u$ as $u \in \mathbf{Z}(P)$. This contradiction shows that $|u^G \cap P| = 3$ for every $1 \ne u \in \mathbf{Z}(P)$. In particular, $|z^G \cap P| = 3$. Now if $\mathbf{Z}(P) \ne P'$, then we can choose $u \in \mathbf{Z}(P) \setminus P'$ and by our previous claim, $|u^G \cap P| = 3$. It follows that u and z are G-conjugate as there is only one class of 2-elements satisfying the previous condition. Again this is a contradiction by using Lemma 2.1(1) and the fact that $z \in \mathbf{Z}(H)$. The proof is now complete. □

We are now ready to prove our first theorem.

Proof of Theorem A. Let G be a finite group. Assume first that G has a normal 2-complement and $d(P) > \frac{1}{2}$ for some Sylow 2-subgroup P of G. By Corollary 2.2(1), $d_2(G) = d(P) > \frac{1}{2}$. Conversely, assume that $d_2(G) > \frac{1}{2}$. Let P be a Sylow 2-subgroup of G. If G has a normal 2-complement, then $d(P) = d_2(G) > \frac{1}{2}$. Thus we only need to show that G has a normal 2-complement. We proceed by induction on $|G|$. Observe that if N is a proper nontrivial normal subgroup of G, then $d_2(N)$ and $d_2(G/N)$ are strictly larger than $\frac{1}{2}$ by Lemma 2.6(2). By induction, both N and G/N have normal 2-complements. Hence if N is of odd order or G/N is a 2-group then G has a normal 2-complement and we are done. Therefore, we may assume that $\mathbf{O}_{2'}(G) = 1$ and $G = \mathbf{O}^2(G)$.

Suppose that $\mathbf{Z}(G)$ is nontrivial. As $\mathbf{O}_{2'}(G) = 1$, $\mathbf{Z}(G)$ must be a 2-group. Now $G/\mathbf{Z}(G)$ has a normal 2-complement, say $K/\mathbf{Z}(G) \trianglelefteq G/\mathbf{Z}(G)$, for some normal subgroup K of G with $\mathbf{Z}(G) \le K$. Hence $\mathbf{Z}(G) \trianglelefteq K \trianglelefteq G$ and G/K is a 2-group. Since $G = \mathbf{O}^2(G)$, we obtain that $G = K$. We now see that $\mathbf{Z}(G)$ is a normal Sylow 2-subgroup of G and thus G has a normal 2-complement by Lemma 2.1(2). So we may assume that $\mathbf{Z}(G) = 1$. Now Lemma 3.1 yields a contradiction. □

We now study the structure of finite groups G with $d_2(G) = \frac{1}{2}$. We first consider the solvable case.

Lemma 3.2. *Let G be a finite solvable group. Suppose that $\boldsymbol{O}_{2'}(G) = 1$ and $G = \boldsymbol{O}^2(G)$. Then $d_2(G) = \frac{1}{2}$ if and only if $G \cong \boldsymbol{A}_4$.*

Proof. If $G \cong \boldsymbol{A}_4$, then $d_2(G) = \frac{1}{2}$ as $k_2(G) = 2$ and $|P| = 4$. Conversely, assume that $d_2(G) = \frac{1}{2}$. We proceed by induction on $|G|$. By Corollary 2.2, we have $\frac{1}{2} = d_2(G) \leq d(P)$. Lemma 2.7 yields that either P is abelian or $|P'| = 2$ and $P/\boldsymbol{Z}(P)$ is elementary abelian of order 4^m. In both cases, $d(P) > \frac{1}{2}$. It follows that G is not a 2-group and so P is noncyclic by Corollary 5.14 in [Isaacs 2008]. As $\boldsymbol{O}_{2'}(G) = 1$, $\boldsymbol{C}_G(\boldsymbol{O}_2(G)) \subseteq \boldsymbol{O}_2(G)$ by [Isaacs 2008, Theorem 3.21], hence $\boldsymbol{Z}(G) \leq P$.

We claim that G/Z satisfies the hypothesis of the lemma for any central subgroup $Z \leq \boldsymbol{Z}(G)$. Clearly, $\boldsymbol{O}^2(G/Z) = G/Z$ since $G = \boldsymbol{O}^2(G)$. Next, assume that $K/Z = \boldsymbol{O}_{2'}(G/Z)$, where $Z \leq K \trianglelefteq G$. Then K has a central Sylow 2-subgroup Z and so by Lemma 2.1(2), K has a normal 2-complement $\boldsymbol{O}_{2'}(K)$. Since $K \trianglelefteq G$, $\boldsymbol{O}_{2'}(K) \leq \boldsymbol{O}_{2'}(G) = 1$. Hence $K = Z$ and so $\boldsymbol{O}_{2'}(G/Z) = 1$.

By Lemma 2.6(2), we have $\frac{1}{2} = d_2(G) \leq d_2(G/Z)d_2(Z) = d_2(G/Z)$. If $d_2(G/Z) > \frac{1}{2}$, then G/Z has a normal 2-complement by Theorem A but this would imply that G/Z is a 2-group and so G is a 2-group, a contradiction. Hence $d_2(G/Z) = \frac{1}{2}$. Therefore, by using induction on $|G|$, if Z is nontrivial, then $G/Z \cong \boldsymbol{A}_4$. We now consider two cases separately, according to whether P is abelian or not.

Case 1: P is abelian. As $\boldsymbol{C}_G(\boldsymbol{O}_2(G)) \subseteq \boldsymbol{O}_2(G)$, we have $P = \boldsymbol{O}_2(G)$ and so $P = \boldsymbol{C}_G(P) \trianglelefteq G$. Clearly, $P \neq \boldsymbol{Z}(G)$, as otherwise G is a 2-group by applying Lemma 2.1(2). Thus $|P : \boldsymbol{Z}(G)| \geq 2$.

Assume first that $\boldsymbol{Z}(G) = 1$. By Lemma 3.1, P is elementary abelian of order 4 and $k_2(G) = 2$. As $P = \boldsymbol{C}_G(P) \trianglelefteq G$, G/P embeds into $\mathrm{GL}_2(2) \cong \boldsymbol{S}_3$. Since G/P is of odd order and nontrivial, $G/P \cong C_3$. It is not hard to see that $G \cong \boldsymbol{A}_4$.

Next, assume that $\boldsymbol{Z}(G)$ is nontrivial. Then $G/\boldsymbol{Z}(G) \cong \boldsymbol{A}_4$. By [Isaacs 2008, Theorem 5.18], $G' \cap \boldsymbol{Z}(G) = G' \cap P \cap \boldsymbol{Z}(G) = 1$. Let R be a Sylow 3-subgroup of G. Then $G = PR$, $|R| = 3$ and R acts nontrivially and coprimely on P; hence $\boldsymbol{Z}(G) = \boldsymbol{C}_P(R)$ and $[P, R] = G' \leq P$. Since R acts coprimely on P, we have $P = [P, R] \times \boldsymbol{C}_P(R) = G' \times \boldsymbol{Z}(G)$. Moreover, $G'R \trianglelefteq G$ and $|G/G'R|$ is a 2-power, so $G = G'R$ forcing $\boldsymbol{Z}(G) = 1$, a contradiction.

Case 2: P is nonabelian. We have $P' \leq \boldsymbol{Z}(P) \leq \boldsymbol{C}_G(\boldsymbol{O}_2(G)) \leq \boldsymbol{O}_2(G)$. Observe that $G/\boldsymbol{O}_{2,2'}(G)$ has an abelian Sylow 2-subgroup, so $G/\boldsymbol{O}_{2,2'}(G)$ has a normal Sylow 2-subgroup by using Hall–Higman Lemma 1.2.3 ([Isaacs 2008, Theorem 3.21]); hence $G = \boldsymbol{O}_{2,2',2,2'}(G)$. (For the definitions of $\boldsymbol{O}_{2,2'}(G)$ and $\boldsymbol{O}_{2,2',2,2'}(G)$, see [Gorenstein 1968, Section 6.3].)

Assume first that $G = \boldsymbol{O}_{2,2'}(G)$. Then $\boldsymbol{C}_G(P) \le P \trianglelefteq G$. It follows that $P' \le \boldsymbol{Z}(G)$ as $|P'| = 2$. Thus $G/P' \cong \boldsymbol{A}_4$ and $|G| = 24$. It is easy to check that $G \cong 2 \cdot \boldsymbol{A}_4 \cong \mathrm{SL}_2(3)$ as $G = \boldsymbol{O}^2(G)$. However, $d_2(\mathrm{SL}_2(3)) = \frac{3}{8} < \frac{1}{2}$.

Assume that $G/\boldsymbol{O}_{2,2'}(G)$ is nontrivial. Let $L = \boldsymbol{O}_{2,2'}(G) \trianglelefteq G$. Since $\boldsymbol{O}_{2'}(L) = 1$, by using Theorem A and Lemma 2.6(2), we see that $d_2(L) = \frac{1}{2}$. But then this forces $d_2(G/L) = 1$. By Corollary 2.2(2), G/L has a normal 2-complement and since $G = \boldsymbol{O}^2(G)$, we deduce that G/L is a $2'$-group, forcing $G = L$, which is a contradiction. This completes our proof. □

Lemma 3.3. *Let G be a finite solvable group. Suppose that $\boldsymbol{O}_{2'}(G) = 1$. If $d_2(G) = \frac{1}{2}$, then $G/\boldsymbol{Z}(G) \cong \boldsymbol{A}_4$ or $\boldsymbol{S}_4$.*

Proof. Suppose that G is a finite solvable group with $d_2(G) = \frac{1}{2}$ and $\boldsymbol{O}_{2'}(G) = 1$. Let $L = \boldsymbol{O}^2(G)$. Then $\boldsymbol{O}_{2'}(L) = 1$ and $L = \boldsymbol{O}^2(L)$. By Lemma 2.6(2), $d_2(L) \ge \frac{1}{2}$. If $d_2(L) > \frac{1}{2}$, then L has a normal 2-complement by Theorem A. However, as $\boldsymbol{O}_{2'}(L) = 1$, L must be a 2-group and hence G is a 2-group with $d_2(G) = d(G) = \frac{1}{2}$, which is impossible by Lemma 2.7. Therefore, $d_2(L) = \frac{1}{2}$. So $L \cong \boldsymbol{A}_4$ by Lemma 3.2.

Let $C = \boldsymbol{C}_G(L) \trianglelefteq G$. As $\boldsymbol{Z}(L) = 1$, we have $C \cap L = 1$. Then $\boldsymbol{A}_4 \cong LC/C \trianglelefteq G/C \le \mathrm{Aut}(\boldsymbol{A}_4) = \boldsymbol{S}_4$. Hence $G/C \cong \boldsymbol{A}_4$ or $\boldsymbol{S}_4$. It remains to show that $C = \boldsymbol{Z}(G)$. Let P be a Sylow 2-subgroup of G.

As $C \times L = CL \trianglelefteq G$, we have $\frac{1}{2} = d_2(G) \le d_2(CL) \le d_2(L)d_2(C) = d_2(C)/2$ and so $d_2(C) = 1$. Thus C has a normal 2-complement and an abelian Sylow 2-subgroup. However, as $\boldsymbol{O}_{2'}(C) \le \boldsymbol{O}_{2'}(G) = 1$, C must be an abelian 2-group and $C \le P$. We also have that $\frac{1}{2} = d_2(G) \le d_2(G/L)d_2(L) = d_2(G/L)/2$, so $d_2(G/L) = 1$ where G/L is a 2-group. It follows that G/L is an abelian 2-group. In particular, $G' \le L$. Thus $[P, C] \subseteq G' \cap C \subseteq L \cap C = 1$, so $[P, C] = 1$. As $[L, C] = 1$ and $G = PL$, we have $C \le \boldsymbol{Z}(G)$. Since $\boldsymbol{Z}(G/C)$ is trivial, we must have $C = \boldsymbol{Z}(G)$ as wanted. □

We next classify all finite nonabelian simple groups S such that $d_2(S) = \frac{1}{2}$.

Lemma 3.4. *Let S be a finite nonabelian simple group. Then $d_2(S) = \frac{1}{2}$ if and only if $S \cong \mathrm{PSL}_2(q)$ with $3 < q \equiv 3, 5 \pmod 8$, where q is a prime power.*

Proof. Let S be a finite nonabelian simple with a Sylow 2-subgroup P. If $S \cong \mathrm{PSL}_2(q)$ with $3 < q \equiv 3, 5 \pmod 8$, then P is elementary abelian of order 4 and S has only one class of involutions so $k_2(S) = 2$ and thus $d_2(S) = \frac{1}{2}$. Conversely, assume that S is a finite nonabelian simple group with $d_2(S) = \frac{1}{2}$. By Lemma 3.1, either P is elementary abelian of order 4 with $k_2(P) = 2$ or P is extra-special of order 2^{1+2m} for some integer $m \ge 1$.

Assume first that P is elementary abelian of order 4. It follows from [Walter 1969, Theorem I] that S is isomorphic to $\mathrm{PSL}_2(2^f)$, where $f \ge 2$, $\mathrm{PSL}_2(q)$ with $3 < q \equiv 3, 5 \pmod 8$ ${}^2\mathrm{G}_2(3^{2n+1})$, where $n \ge 1$, or the first Janko group J_1. Since $|P| = 4$, we deduce that $S \cong \mathrm{PSL}_2(q)$ with $3 < q \equiv 3, 5 \pmod 8$. Note $\mathrm{PSL}_2(4) \cong \mathrm{PSL}_2(5)$.

Assume now that P is extra-special of order 2^{1+2m}, $m \geq 1$. In this case, P is nilpotent of class 2. It follows from the Main Theorem of [Gilman and Gorenstein 1975a; 1975b] that S is isomorphic to one of the groups $\mathrm{PSL}_2(q)$ with $q \equiv 7, 9 \pmod{16}$, $\boldsymbol{A}_7$, $\mathrm{Sz}(2^n)$, $\mathrm{PSU}_3(2^n)$, $\mathrm{PSL}_3(2^n)$ or $\mathrm{PSp}_4(2^n)$ with $n \geq 2$. However, except for the first two groups, the centers of the Sylow 2-subgroups of the remaining simple groups have order at least 4. For $\boldsymbol{A}_7$, we can check that $k_2(\boldsymbol{A}_7) = 3$ so $d_2(\boldsymbol{A}_7) = \frac{3}{8}$ as a Sylow 2-subgroup of $\boldsymbol{A}_7$ is isomorphic to D_8. Similarly, the Sylow 2-subgroup of $S = \mathrm{PSL}_2(q)$ with $q \equiv 7, 9 \pmod{16}$ is also isomorphic to D_8. Again, except for the identity, S has two nontrivial classes of 2-elements, one consisting of all involutions in S and another consisting of elements of order 4. Thus these cases cannot occur. □

For a finite group G, we denote by Sol(G) the solvable radical of G, that is, the largest solvable normal subgroup of G.

Lemma 3.5. *Let G be a finite group. Suppose that* Sol(G) $= 1$ *and* $d_2(G) = \frac{1}{2}$. *Then G is a finite almost simple group.*

Proof. Let M be a minimal normal subgroup of G. As G has a trivial solvable radical, $M \cong S^k$, where S is a nonabelian simple group and $k \geq 1$ is an integer. By Lemma 2.6(2), we have $\frac{1}{2} = d_2(G) \leq d_2(G/M)d_2(M) \leq d_2(M)$. By applying this lemma repeatedly, we have $\frac{1}{2} \leq d_2(M) \leq d_2(S)^k$. By Lemma 3.1, $d_2(S) \leq \frac{1}{2}$; so $\frac{1}{2} \leq d_2(S)^k \leq (\frac{1}{2})^k$, forcing $k = 1$ and $d_2(S) = \frac{1}{2}$.

Let $C = \boldsymbol{C}_G(M)$. Then $C \trianglelefteq G$ and $CM = C \times M \trianglelefteq G$. By Lemma 2.6(2),

$$\tfrac{1}{2} = d_2(G) \leq d_2(G/MC)d_2(MC) \leq d_2(MC) \leq d_2(M)d_2(C) = d_2(C)/2.$$

Hence $d_2(C) = 1$ and so C is solvable by Corollary 2.2(2) and the Feit–Thompson theorem. Since Sol(G) $= 1$ and C is a solvable normal subgroup of G, we must have $C = 1$ so G is almost simple with simple socle M. □

Lemma 3.6. *Let G be a finite perfect group. Suppose that* $\boldsymbol{O}_{2'}(G) = 1$ *and* $d_2(G) = \frac{1}{2}$. *Then* $G \cong \mathrm{PSL}_2(q)$ *with* $3 < q \equiv 3, 5 \pmod 8$, *where q is a prime power.*

Proof. Let U be the solvable radical of G. Then G/U is nonsolvable. Since $\frac{1}{2} = d_2(G) \leq d_2(U)d_2(G/U)$ by Lemma 2.6, both $d_2(U)$ and $d_2(G/U)$ are at least $\frac{1}{2}$. By Theorem A, $d_2(G/U) = \frac{1}{2}$ as otherwise G/U is solvable. By Lemmas 3.4 and 3.5 and the fact that G is perfect, $G/U \cong \mathrm{PSL}_2(q)$, where $q \equiv 3, 5 \pmod 8$. We have $d_2(U) = 1$ and since $\boldsymbol{O}_{2'}(G) = 1$, U is an abelian 2-group. We will show that G is nonabelian simple by induction on $|G|$.

If $U = 1$, then G is simple and we are done. Assume that U is nontrivial. Assume first that $Z := \boldsymbol{Z}(G)$ is nontrivial. Then Z must be a 2-group. Consider the quotient group G/Z. Observe that G/Z is perfect, $d_2(G/Z) = \frac{1}{2}$ and $\boldsymbol{O}_{2'}(G/Z) = 1$. Since $|G/Z| < |G|$, the inductive hypothesis implies that G/Z is nonabelian simple and thus $Z = U$. It follows that $G \cong \mathrm{SL}_2(q)$, the only Schur cover of $\mathrm{PSL}_2(q)$ with $q \equiv 3, 5 \pmod 8$. However, it is easy to see that $\mathrm{SL}_2(q)$ has only two classes of nontrivial 2-elements and the Sylow 2-subgroup of $\mathrm{SL}_2(q)$ with $q \equiv 3, 5 \pmod 8$ has order 8, so $d_2(G) = \frac{3}{8} < \frac{1}{2}$, which is a contradiction. Hence we may assume that $\boldsymbol{Z}(G) = 1$. Since U is a normal abelian subgroup of G, we have $U \le \boldsymbol{C}_G(U) \trianglelefteq G$. Since U is not central in G and G/U is nonabelian simple, we must have that $U = \boldsymbol{C}_G(U)$.

Let $P \in \mathrm{Syl}_2(G)$. Note that the hypothesis of Lemma 3.1 holds for G, that is, $\boldsymbol{O}_{2'}(G) = \boldsymbol{Z}(G) = 1$, $|P| > 1$ and that $d_2(G) = \frac{1}{2}$. We claim that $k_2(G) > 2$. Assume by contradiction that $k_2(G) = 2$. Then P is elementary abelian of order 4 by Lemma 3.1. However the Sylow 2-subgroup of $\mathrm{PSL}_2(q)$ with $q \equiv 3, 5 \pmod 8$ has order 4. So $U = 1$, which is a contradiction. Therefore $k_2(G) > 2$ and so part (2) of Lemma 3.1 holds. Obviously $|U| \ge 4$ and $P' = \langle z \rangle = \boldsymbol{Z}(P) < U$. If $z^G = U \setminus \{1\}$, then U is elementary abelian of order 4 and thus G/U embeds into $\mathrm{GL}_2(2)$ which is impossible. Thus there exists $1 \ne y \in U \setminus z^G$ and so $|y^G \cap P| = 2$. Since $y \in U \trianglelefteq G$, we have $y^G \subseteq U \le P$, so $|y^G \cap P| = |y^G| = 2$ which implies that $U \le \boldsymbol{C}_G(y) < G$ and $|G : \boldsymbol{C}_G(y)| = 2$. Therefore, G/U has a subgroup of index 2 which is impossible as G/U is nonabelian simple. □

Proof of Theorem B. Let G be a finite group and assume that $d_2(G) = \frac{1}{2}$ and $\boldsymbol{O}_{2'}(G) = 1$. If G is solvable, then $G/\boldsymbol{Z}(G) \cong \boldsymbol{A}_4$ or $\boldsymbol{S}_4$ by Lemma 3.3. So part (1) of the theorem holds. Assume that G is nonsolvable. Let L be the last term of the derived series of G. By Theorem A and Lemma 2.6(2), $d_2(L) = \frac{1}{2}$. Moreover L is perfect and $\boldsymbol{O}_{2'}(L) = 1$. By Lemma 3.6, $L \cong S$ where $S = \mathrm{PSL}_2(q)$ $q \equiv 3, 5 \pmod 8$. Write $q = p^f$, where p is a prime and $f \ge 1$ is an integer. We see that f must be odd and thus $\mathrm{Out}(S) = C_2 \times C_f$.

Let $C = \boldsymbol{C}_G(L)$. Then $C \trianglelefteq G$, $C \cap L = 1$ and G/C is an almost simple group with socle isomorphic to S. Since $d_2(L) = \frac{1}{2}$, we see that $d_2(C) = 1$ and since $\boldsymbol{O}_{2'}(G) = 1$, C is a normal abelian 2-subgroup of G. We also have that

$$\tfrac{1}{2} = d_2(G) \le d_2(G/L)d_2(L) = d_2(G/L)/2$$

so $d_2(G/L) = 1$ and so G/L has a normal 2-complement W/L and an abelian Sylow 2-subgroup PL/L by Corollary 2.2, where P is any Sylow 2-subgroup of G containing C. Since CL/L and W/L are normal subgroups of G/L and have coprime orders, we deduce that $[C, W] \le L$. As $C \trianglelefteq G$, we have $[C, W] \le L \cap C = 1$. Thus $[C, W] = 1$. On the other hand, PL/L is abelian, thus $[C, P] \le L$. With the same reasoning, we have $[C, P] \le L \cap C = 1$. Since $G = PW$, we obtain that $[C, G] = 1$. In particular, $C \le \boldsymbol{Z}(G)$ and since G/C is almost simple, we must have that $C = \boldsymbol{Z}(G)$. Therefore, we have shown that $G/\boldsymbol{Z}(G)$ is almost simple with socle S as required. □

We will need the following result for our proof of Theorem C.

Lemma 3.7. *Let G be a finite group of odd order and let σ be a nonempty set of primes. If $d_\sigma(G) \geq \frac{1}{2}$, then G has a normal σ-complement and an abelian Hall σ-subgroup.*

Proof. By the Feit–Thompson theorem, we know G is solvable. By Lemma 2.6(2), if $N \trianglelefteq G$, then $d_\sigma(N) \geq \frac{1}{2}$ and $d_\sigma(G/N) \geq \frac{1}{2}$. Assume that G has a normal σ-complement K. Let H be a Hall σ-subgroup of G. We claim that H is abelian. As $G/K \cong H$, we have $\frac{1}{2} \leq d_\sigma(H) = d(H)$, where the last equality holds as H is a σ-group. Thus $d(H) \geq \frac{1}{2}$ where H is a group of odd order. By Lemma 2.7, H must be abelian as wanted. Therefore, it suffices to show that G has a normal σ-complement. We will prove this claim by induction on $|G|$.

Let N be a minimal normal subgroup of G. Then N is an elementary abelian p-subgroup for some odd prime p. As $d_\sigma(G/N) \geq \frac{1}{2}$, by induction on $|G|$, G/N has a normal σ-complement, say M/N. If $p \notin \sigma$, then M is also a normal σ-complement of G, and we are done. Thus we may assume that $\boldsymbol{O}_{\sigma'}(G) = 1$ and $p \in \sigma$. We have $M \trianglelefteq G$ and $d_\sigma(M) \geq \frac{1}{2}$. Therefore, by induction again, M has a normal σ-complement whenever $M < G$; but then this would imply that M is a σ-subgroup since $\boldsymbol{O}_{\sigma'}(M) \subseteq \boldsymbol{O}_{\sigma'}(G) = 1$ and hence G is a σ-group. So, we can assume $M = G$, hence G/N is a σ'-group.

Since G/N is solvable, let T/N be a maximal normal subgroup of G/N of prime index $r \notin \sigma$. Since $T \trianglelefteq G$, we have $d_\sigma(T) \geq \frac{1}{2}$ and again by induction, T has a normal σ-complement which implies that $T = N$. Thus N is a maximal normal subgroup of G and $|G/N| = r$ is a prime different from p.

If $\boldsymbol{C}_G(x) = G$ for some $1 \neq x \in N$, then $\langle x \rangle = N \leq \boldsymbol{Z}(G)$ and $G \cong C_p \times C_r$ by Lemma 2.1(2), which is a contradiction as $\boldsymbol{O}_{\sigma'}(G) = 1$. So, we may assume that $\boldsymbol{C}_G(x) < G$ for all $1 \neq x \in N$. Since N is maximal in G, $\boldsymbol{C}_G(x) = N$ for every $1 \neq x \in N$. Thus G is a Frobenius group with Frobenius kernel N and Frobenius complement isomorphic to C_r. Set $|N| = p^k$ for some integer $k \geq 1$. We can see that $\sigma = \{p\}$ and that $k_p(G) = (p^k - 1)/r + 1$ and so $d_p(G) = 1/r + (r-1)/(rp^k) \geq \frac{1}{2}$. Notice that $r \neq p \geq 3$ and $r \mid p^k - 1$. We consider the following cases:

(1) $r = 3$ and $p \geq 5$. In this case, we have

$$d_p(G) \leq \tfrac{1}{3} + \tfrac{2}{3} \cdot \tfrac{1}{5} = \tfrac{7}{15} < \tfrac{1}{2}.$$

(2) $r \geq 5$ and $p = 3$. Since $r > p$, $k \geq 2$. We have

$$d_p(G) \leq \tfrac{1}{5} + \tfrac{1}{9} = \tfrac{14}{45} < \tfrac{1}{2}.$$

(3) $r \geq 5$ and $p \geq 5$. Clearly, we have

$$d_p(G) < \tfrac{1}{5} + \tfrac{1}{5} = \tfrac{2}{5} < \tfrac{1}{2}.$$

Thus we have shown that G has a normal σ-complement as wanted. □

We are now ready to prove Theorem C.

Proof of Theorem C. Let G be a finite group and let π be a set of primes containing 2 and let $\sigma = \pi \setminus \{2\}$. Suppose $d_\pi(G) > \frac{1}{2}$. By Lemma 2.6(1), $\frac{1}{2} < d_\pi(G) \le d_2(G)$ and thus by Theorem A, G has a normal 2-complement K and by Lemma 2.6(2), we have

$$\tfrac{1}{2} < d_\pi(G) \le d_\pi(K) d_\pi(G/K) \le d_\pi(K) = d_\sigma(K).$$

By Lemma 3.7 , K has a normal σ-complement, say N, and an abelian Hall σ-subgroup T. It follows that $G = PTN$, where N is also a normal π-complement of G. □

4. Conjugacy classes of p-elements with p odd

We now consider odd primes. We start with the following easy result.

Lemma 4.1. *Let p be an odd prime. Let G be a finite group and let P be a Sylow p-subgroup of G. If $d_p(G) \ge (p+1)/2p$, then P is abelian.*

Proof. Let G be a finite group such that $d_p(G) \ge (p+1)/2p$. By Corollary 2.2, we have $d_p(G) \le d(P)$ which implies that $d(P) \ge (p+1)/2p$. If P is abelian, then we are done. So, assume that P is nonabelian. By Lemma 2.6(3), we have $d(P) < (p+1)/p^2$. Since p is odd, we can check that $(p+1)/2p > (p+1)/p^2$ and so $d(P) < (p+1)/p^2 < (p+1)/2p \le d(P)$, which is a contradiction. □

Proof of Theorem D. Let p be an odd prime. Let G be a finite group. Assume that G has a normal p-complement and an abelian Sylow p-subgroup P. By Corollary 2.2(2), we have $d_p(G) = d(P) = 1 > (p+1)/2p$. Conversely, assume that $d_p(G) > (p+1)/2p$. By Lemma 4.1, P is abelian. It remains to show that G has a normal p-complement. We proceed by using induction on $|G|$.

We first claim that $\boldsymbol{O}_{p'}(G) = 1$. Assume by contradiction that $\boldsymbol{O}_{p'}(G)$ is nontrivial. By Lemma 2.6(2), $d_p(G) \le d_p(G/\boldsymbol{O}_{p'}(G)) d_p(\boldsymbol{O}_{p'}(G)) \le d_p(G/\boldsymbol{O}_{p'}(G))$, so by induction $G/\boldsymbol{O}_{p'}(G)$ has a normal p-complement; hence G will have a normal p-complement. Thus we may assume that $\boldsymbol{O}_{p'}(G) = 1$.

We next claim that $G = \boldsymbol{O}^p(G)$. Indeed, if $N = \boldsymbol{O}^p(G)$ is a proper subgroup of G, then $(p+1)/2p < d_p(G) \le d_p(N)$; thus by induction again, N has a normal p-complement $\boldsymbol{O}_{p'}(N)$. Clearly, this is also a normal p-complement of G.

We now show that G is p-solvable. In fact, suppose that G is not p-solvable and let M/N be a nonabelian chief factor of G with p dividing $|M/N|$. There exists a nonabelian simple group S and an integer $k \ge 1$ such that $M/N \cong S^k$. By applying Lemma 2.6(2) repeatedly, we have $(p+1)/2p < d_p(S)^k \le d_p(S)$. (Note that p divides $|S|$.) Let $T \in \mathrm{Syl}_p(S)$ and let $H = \boldsymbol{N}_S(T)$. Clearly T is abelian, so by Lemma 2.1(1), H controls S-fusion in T. Thus $x^S \cap T = x^H \subseteq T$ for every $x \in T$. Since S is nonabelian simple, $\boldsymbol{Z}_p^*(S) = 1$. Now Lemmas 2.3 and 2.5 together with

Glauberman's Z_p^*-theorem imply that $|x^S \cap T| \geq 2$ for all $1 \neq x \in T$. It follows that $|T| - 1 \geq 2(k_p(S) - 1)$. This implies that $k_p(S) \leq (|T| + 1)/2$ and hence

$$(p+1)/2p < d_p(S) \leq (|T|+1)/(2|T|) \leq (p+1)/2p$$

as $|T| \geq p$. This contradiction shows that G is p-solvable.

By Hall–Higman Lemma 1.2.3 ([Isaacs 2008, Lemma 3.21]) and the fact that P is abelian, we have $P \leq \boldsymbol{C}_G(\boldsymbol{O}_p(G)) \leq \boldsymbol{O}_p(G)$, so

$$P = \boldsymbol{O}_p(G) \trianglelefteq G.$$

Let P/N be a chief factor of G. Assume that N is nontrivial. Then

$$(p+1)/2p < d_p(G) \leq d_p(G/N)$$

and so by induction G/N has a normal p-complement K/N. However, as $G = \boldsymbol{O}^p(G)$, $G = K$ which is impossible. So, we can assume that P is an elementary abelian minimal normal p-subgroup of G.

If $|x^G| = 1$ for some $1 \neq x \in P$, then $x \in \boldsymbol{Z}(G) \cap P$ which forces $P = \langle x \rangle \subseteq \boldsymbol{Z}(G)$. In this case, G has a normal p-complement by Lemma 2.1(2). Hence we can also assume that $|x^G| \geq 2$ for all $1 \neq x \in P$ whence $k_p(G) \leq (|P| + 1)/2$. Since $d_p(G) > (p+1)/2p$,

$$|P|(p+1)/2p < (|P|+1)/2.$$

However, this inequality cannot occur as $|P| \geq p$. □

Proof of Theorem E. Let π be a nonempty set of odd primes and let p be the smallest member in π. Let G be a finite group with $d_\pi(G) > (p+1)/2p$. For every $r \in \pi$, we see that

$$(r+1)/2r \leq (p+1)/2p < d_\pi(G) \leq d_r(G)$$

by Lemma 2.6(1), so $d_r(G) > (r+1)/2r$. By Theorem D, G has a normal r-complement and an abelian Sylow r-subgroup. It follows that G has a normal π-complement $N = \boldsymbol{O}_{\pi'}(G) \trianglelefteq G$ and G is π-solvable. By [Isaacs 2008, Theorem 3.20], G has a Hall π-subgroup H. Clearly $G = HN$ and $G/N \cong H$. Since

$$(p+1)/2p < d_\pi(G) \leq d_\pi(H)d_\pi(N) \leq d_\pi(H) = d(H),$$

we deduce that

$$d(H) > (p+1)/2p > 1/p \geq 1/r$$

and so by Lemma 2.6(4), H has a normal Sylow r-subgroup. It follows that H is nilpotent and thus H is abelian. □

Proof of Theorem F. Let G be a finite group with a Sylow p-subgroup P, where p is an odd prime. Suppose that $d_p(G) = (p+1)/2p$ and $\boldsymbol{O}_{p'}(G) = 1$. By Lemma 4.1,

P is abelian and thus by Lemma 2.1(1), $d_p(\boldsymbol{N}_G(P)) = d_p(G) = (p+1)/2p$. By Theorem 7.4.4 in [Gorenstein 1968], we have $P = P \cap N' \times P \cap \boldsymbol{Z}(N)$, where $N = \boldsymbol{N}_G(P)$. Set $Z = P \cap \boldsymbol{Z}(N)$ and $U = P \cap N'$. We have that $k_p(N) \le (|P|+|Z|)/2$ as for any $x \in P \setminus Z$, $|x^N| \ge 2$. It follows that

$$(p+1)/2p \le (|P|+|Z|)/2|P|$$

and thus $|P : Z| \le p$. Clearly, $|P : Z| > 1$ as otherwise $P \subseteq \boldsymbol{Z}(N)$ and thus G has a normal p-complement which forces $G = P$ since $\boldsymbol{O}_{p'}(G) = 1$. But then $d_p(G) = 1$, a contradiction. Thus $|P : Z| = p$; hence $|U| = p$ and $|Z| = |P|/p$. Moreover $|x^N| = 2$ for every $x \in P \setminus Z$. Set $U = \langle y \rangle$. Then $\boldsymbol{C}_G(P) = \boldsymbol{C}_N(y) \trianglelefteq N$. As $|y^N| = 2$, we have $|N : \boldsymbol{C}_G(P)| = 2$. Since $U = P \cap N'$ has order p, we see that $U = [P, N]$. Furthermore, by Theorem 7.4.4 in [Gorenstein 1968], $P \cap G' = P \cap N' = U$ is cyclic of order p and by Theorem 5.18 in [Isaacs 2008], $G' \cap Z = 1$.

Let $\boldsymbol{F}^*(G)$ be the generalized Fitting subgroup of G. Then $\boldsymbol{F}^*(G) = \boldsymbol{F}(G)\boldsymbol{E}(G)$ is the central product of the Fitting subgroup $\boldsymbol{F}(G)$ and the layer $\boldsymbol{E}(G)$ of G, which is the product of all components of G, that is, subnormal quasisimple subgroups of G. Bender's theorem ([Isaacs 2008, Theorem 9.8]) says that $\boldsymbol{C}_G(\boldsymbol{F}^*(G)) \subseteq \boldsymbol{F}^*(G)$.

Assume first that $\boldsymbol{E}(G) = 1$. Then $\boldsymbol{F}^*(G) = \boldsymbol{F}(G)$ is a p-group since $\boldsymbol{O}_{p'}(G) = 1$. As $\boldsymbol{C}_G(\boldsymbol{F}(G)) \subseteq \boldsymbol{F}(G)$ and P is abelian, $P = \boldsymbol{F}(G)$ and thus $P = \boldsymbol{C}_G(P) \trianglelefteq G$. It follows that $|G : P| = 2$ and $G' = [G, P] = U$ is cyclic of order p; moreover $Z = \boldsymbol{Z}(G) \cap P = \boldsymbol{Z}(G)$ as P is self-centralizing. Now G/G' is an abelian group of order $2|Z|$. Hence G/G' has a normal Sylow 2-subgroup A/G' and a normal Sylow p-subgroup $P/G' = ZG'/G'$, so $G/G' = A/G' \times ZG'/G'$ which implies that $G = Z \times A$, where A is a nonabelian group of order $2p$ and it has a normal cyclic Sylow p-subgroup of order p. It is easy to see that $A \cong \mathrm{D}_{2p}$, the dihedral group of order $2p$.

Assume now that $E := \boldsymbol{E}(G)$ is nontrivial. Since $G' \cap P$ is cyclic of order p, the center of E is either trivial or cyclic of order p. If $|\boldsymbol{Z}(E)| = p$, then $E/\boldsymbol{Z}(E)$ is a p'-group which is impossible by Corollary 5.4 in [Isaacs 2008]. Thus $\boldsymbol{Z}(E) = 1$. Hence E has a Sylow p-subgroup of order p which forces E to be a nonabelian simple group. Now we have that $\boldsymbol{F}^*(G) = E \times F$ where $F = \boldsymbol{F}(G) \le P$ is a p-subgroup. Since $E \cap P \le G' \cap P = U$ is of order p, we have $U = E \cap P = G' \cap P$. Therefore, $F \cap G' \le F \cap G' \cap P = E \cap P \cap F = 1$, so $F \cap G' = 1$, whence $F \le \boldsymbol{Z}(G)$. Hence $\boldsymbol{C}_G(E) = F = \boldsymbol{Z}(G)$ and so G/F is an almost simple group with socle isomorphic to E. Since E has a cyclic Sylow p-subgroup of order p, we deduce from Lemma 2.3 in [Tong-Viet 2018] that $|\mathrm{Out}(E)|$ is prime to p. In particular, G/EF is a solvable p'-group. Thus G/E has a central Sylow p-subgroup $EF/E \cong F$. By Lemma 2.1(2), G/E has a normal p-complement A/E and $G/E = EF/E \times A/E$. Since $E \cap F = 1$, we have $G = A \times F$ and so A is almost simple with socle E. □

Acknowledgments

The author is grateful to the anonymous referees for many useful comments and suggestions.

References

[Fulman and Guralnick 2012] J. Fulman and R. Guralnick, "Bounds on the number and sizes of conjugacy classes in finite Chevalley groups with applications to derangements", *Trans. Amer. Math. Soc.* **364**:6 (2012), 3023–3070. MR Zbl

[Gilman and Gorenstein 1975a] R. Gilman and D. Gorenstein, "Finite groups with Sylow 2-subgroups of class two, I", *Trans. Amer. Math. Soc.* **207** (1975), 1–101. MR Zbl

[Gilman and Gorenstein 1975b] R. Gilman and D. Gorenstein, "Finite groups with Sylow 2-subgroups of class two, II", *Trans. Amer. Math. Soc.* **207** (1975), 103–126. MR Zbl

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

[Gorenstein 1968] D. Gorenstein, *Finite groups*, Harper & Row, New York, 1968. MR Zbl

[Guralnick and Robinson 1993] R. M. Guralnick and G. R. Robinson, "On extensions of the Baer–Suzuki theorem", *Israel J. Math.* **82**:1-3 (1993), 281–297. MR Zbl

[Guralnick and Robinson 2006] R. M. Guralnick and G. R. Robinson, "On the commuting probability in finite groups", *J. Algebra* **300**:2 (2006), 509–528. MR Zbl

[Gustafson 1973] W. H. Gustafson, "What is the probability that two group elements commute?", *Amer. Math. Monthly* **80** (1973), 1031–1034. MR Zbl

[Isaacs 2008] I. M. Isaacs, *Finite group theory*, Grad. Studies in Math. **92**, Amer. Math. Soc., Providence, RI, 2008. MR Zbl

[Külshammer et al. 2014] B. Külshammer, G. Navarro, B. Sambale, and P. H. Tiep, "Finite groups with two conjugacy classes of p-elements and related questions for p-blocks", *Bull. Lond. Math. Soc.* **46**:2 (2014), 305–314. MR Zbl

[Lescot 1995] P. Lescot, "Isoclinism classes and commutativity degrees of finite groups", *J. Algebra* **177**:3 (1995), 847–869. MR Zbl

[Lescot 2001] P. Lescot, "Central extensions and commutativity degree", *Comm. Algebra* **29**:10 (2001), 4451–4460. MR Zbl

[Maróti and Nguyen 2014] A. Maróti and H. N. Nguyen, "On the number of conjugacy classes of π-elements in finite groups", *Arch. Math.* (*Basel*) **102**:2 (2014), 101–108. MR Zbl

[Tong-Viet 2018] H. P. Tong-Viet, "Brauer characters and normal Sylow p-subgroups", *J. Algebra* **503** (2018), 265–276. MR Zbl

[Walter 1969] J. H. Walter, "The characterization of finite groups with abelian Sylow 2-subgroups", *Ann. of Math.* (2) **89** (1969), 405–514. MR Zbl

Received March 1, 2020. Revised May 22, 2020.

HUNG P. TONG-VIET
DEPARTMENT OF MATHEMATICAL SCIENCES
BINGHAMTON UNIVERSITY
BINGHAMTON, NY 13902-6000
UNITED STATES
tongviet@math.binghamton.edu

PACIFIC JOURNAL OF MATHEMATICS
Vol. 308, No. 1, 2020
https://doi.org/10.2140/pjm.2020.308.223

NOT EVEN KHOVANOV HOMOLOGY

PEDRO VAZ

We construct a supercategory that can be seen as a skew version of (thickened) KLR algebras for the type A quiver. We use our supercategory to construct homological invariants of tangles and show that for every link our invariant gives a link homology theory supercategorifying the Jones polynomial. Our homology is distinct from even Khovanov homology and we present evidence supporting the conjecture that it is isomorphic to odd Khovanov homology. We also show that cyclotomic quotients of our supercategory give supercategorifications of irreducible finite-dimensional representations of $\mathfrak{gl}_n$ of level 2.

1. Introduction

After the appearance of odd Khovanov homology in [Ozsváth et al. 2013] there has been a certain interest in odd categorified structures and supercategorification (see, for example, [Lauda and Egilmez 2018; Ellis et al. 2014; Ellis and Lauda 2016; Ellis and Qi 2016; Kang et al. 2013; 2014; Lauda and Russell 2014; Naisse and Vaz 2018]). In contrast to (even) Khovanov homology, odd Khovanov homology has an anticommutative feature. Both theories categorify the Jones polynomial and both agree modulo 2, but they are intrinsically distinct (see [Shumakovitch 2011] for a study of the properties of odd Khovanov homology and a comparison with even Khovanov homology).

A construction of odd Khovanov homology using higher representation theory is still missing. In the case of even Khovanov homology this question was solved in [Webster 2017] using categorification of tensor products and the WRT invariant and in [Lauda et al. 2015] using categorical Howe duality.

In this paper we construct a supercategorification of the Jones invariant for tangles using higher representation theory. In particular, we define a supercategory in the spirit of Khovanov and Lauda's diagrammatics that can be seen as a superalgebra version of KLR algebras [Khovanov and Lauda 2009; Rouquier 2008] of level 2 for the A_n quiver. We present our supercategory in the form of a graphical calculus reminiscent of the thick calculus for categorified $\mathfrak{sl}_2$ [Khovanov et al. 2012] and $\mathfrak{sl}_n$

MSC2010: primary 81R50; secondary 17B37, 18G60, 57M25.
Keywords: odd Khovanov homology, categorification, higher representation theory, KLR algebras.

[Stošić 2019] (see also [Ellis et al. 2014] for a thick calculus for the odd nilHecke algebra). Our supercategory admits cyclotomic quotients that supercategorify irreducibles of $U_q(\mathfrak{gl}_k)$ of level 2.

We use cyclotomic quotients of our supercategories as input to Tubbenhauer's approach [2014] to Khovanov–Rozansky homologies. It is based on q-Howe duality and uses only the lower half of the quantum group $U_q(\mathfrak{gl}_k)$ to produce an invariant of tangles. In our case we obtain an invariant that shares several similarities with odd Khovanov homology when restricted to links. For example, it decomposes as a direct sum of two copies of a reduced homology and it produces chronological Frobenius algebras, analogous to the ones that can be extracted from [Ozsváth et al. 2013] (see [Putyra 2014a] for explanations). Both theories coincide over $\mathbb{Z}/2\mathbb{Z}$. We also give computational evidence that our invariant is distinct from even Khovanov homology and that support the conjecture that for every link L it coincides with the odd Khovanov homology of L.

2. The supercategory $\mathfrak{R}$

2A. ***The supercategory $\mathfrak{R}(\nu)$.*** We follow [Brundan and Ellis 2017] regarding supercategories. For objects X, Y in a supercategory $\mathcal{C}$ we write $\mathrm{Hom}^0_{\mathcal{C}}(X, Y)$ (resp. $\mathrm{Hom}^1_{\mathcal{C}}(X, Y)$) for its space of even (resp. odd) morphisms and we write $p(f)$ for the parity of $f \in \mathrm{Hom}^i_{\mathcal{C}}(X, Y)$. If $\mathcal{C}$ has additionally a $\mathbb{Z}$-grading we denote by $q^s X$ a grading shift up of X by s units and we consider only morphisms that preserve the $\mathbb{Z}$-grading. In this case we write $\mathrm{Hom}_{\mathcal{C}}(X, Y) = \bigoplus_{s\in\mathbb{Z}} \mathrm{Hom}_{\mathcal{C}}(X, q^s Y)$. We follow the grading conventions in [Lauda et al. 2015], which are aligned with the tradition in link homology. This means that a map of degree s from X to Y yields a degree zero map from X to $q^s Y$.

Fix a unital ring $\Bbbk$. Let $\alpha_1, \dots, \alpha_n$ denote the simple roots of $\mathfrak{sl}_n$ and $\langle -, - \rangle$ their inner product: $\langle \alpha_i, \alpha_i \rangle = 2$, $\langle \alpha_i, \alpha_{i\pm1} \rangle = -1$, and $\langle \alpha_i, \alpha_j \rangle = 0$ otherwise. Fix also a choice of scalars Q consisting of $r_i, t_{ij} \in \Bbbk^\times$ for all $i, j \in I := \{1, \dots, n\}$, such that $t_{ii} = 1$ and $t_{ij} = t_{ji}$ when $|i - j| \neq 1$. Let also p_{ij} be defined by $p_{ii} = p_{i+1,i} = 1$ and otherwise $p_{ij} = 0$.

For each $\nu = \sum_{i\in I} \nu_i . i \in \mathbb{N}_0[I]$, we consider the set of (colored) sequences of ν,

$$\mathrm{CSeq}(\nu) := \Big\{ i_1^{(\varepsilon_1)} \cdots i_r^{(\varepsilon_r)} \mid \varepsilon_s \in \{1, 2\},\ \sum_s \varepsilon_s i_s = \nu \Big\}.$$

By convention we write simply i_s for $i_s^{(1)}$. Two sequences $\boldsymbol{i} \in \mathrm{CSeq}(\nu)$ and $\boldsymbol{j} \in \mathrm{CSeq}(\nu')$ can be concatenated into a sequence $\boldsymbol{i}\boldsymbol{j}$ in $\mathrm{CSeq}(\nu + \nu')$.

Definition 2.1. The supercategory $\mathfrak{R}(\nu)$ is defined by the following data:

(a) The objects of $\mathfrak{R}(\nu)$ are finite formal sums of grading shifts of elements of $\mathrm{CSeq}(\nu)$.

(b) The morphism space $\mathrm{Hom}_{\mathfrak{R}(\nu)}(\boldsymbol{i},\boldsymbol{j})$ from $\boldsymbol{i}$ to $\boldsymbol{j}$ is the $\mathbb{Z}$-graded $\Bbbk$-supervector space generated by vertical juxtaposition and horizontal juxtaposition of the diagrams below. Composition consists of vertical concatenation of diagrams. By convention we read diagrams from bottom to top and so, ab consists of stacking the diagram for a atop the one for b. Diagrams are equipped with a Morse function that keeps trace of the relative height of the generators. We consider isotopy classes of such diagrams that do not change the relative height of generators.

Generators.

- Simple and double *identities*

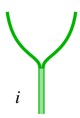
 $\in \mathrm{Hom}^0_{\mathfrak{R}(\nu)}(i,i)$, i

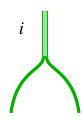
 $\in \mathrm{Hom}^0_{\mathfrak{R}(\nu)}(i^{(2)},i^{(2)})$, i

- *dots*

 $\in \mathrm{Hom}^1_{\mathfrak{R}(\nu)}(i,q^2 i)$, i

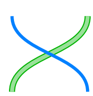
- *splitters*

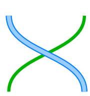
 $\in \mathrm{Hom}^1_{\mathfrak{R}(\nu)}(i^{(2)},q^{-1}ii)$, i

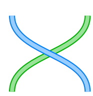
i  $\in \mathrm{Hom}^0_{\mathfrak{R}(\nu)}(ii,q^{-1}i^{(2)})$,

- and *crossings*

 $\in \mathrm{Hom}^{p_{ij}}_{\mathfrak{R}(\nu)}(ij,q^{-\langle\alpha_i,\alpha_j\rangle}ji)$, i j

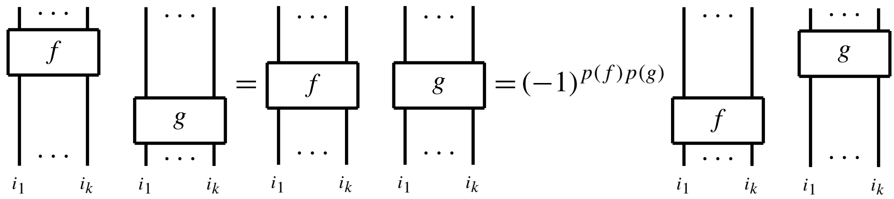
 $\in \mathrm{Hom}^0_{\mathfrak{R}(\nu)}(i^{(2)}j,q^{-2\langle\alpha_i,\alpha_j\rangle}ji^{(2)})$, i j

 $\in \mathrm{Hom}^0_{\mathfrak{R}(\nu)}(ij^{(2)},q^{-2\langle\alpha_i,\alpha_j\rangle}j^{(2)}i)$, i j

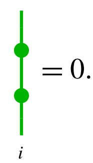
 $\in \mathrm{Hom}^0_{\mathfrak{R}(\nu)}\big(i^{(2)}j^{(2)},q^{-4\langle\alpha_i,\alpha_j\rangle}j^{(2)}i^{(2)}\big)$. i j

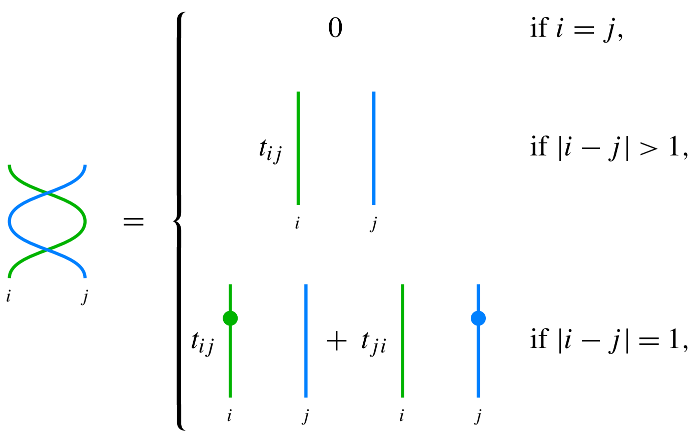
Relations. Morphisms are subject to the local relations (1) to (14) below.

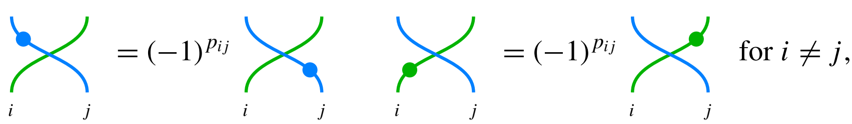
- For all f, g:

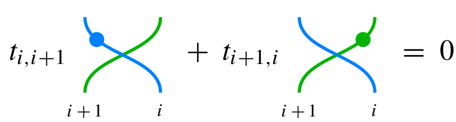
(1) 

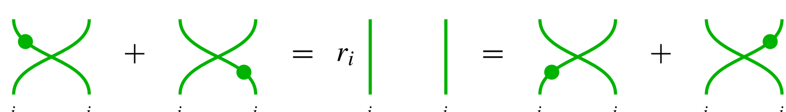
- For all $i, j, k \in I$:

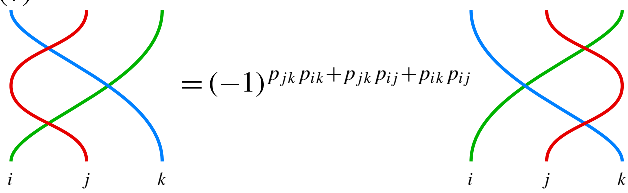
(2) 

(3) 

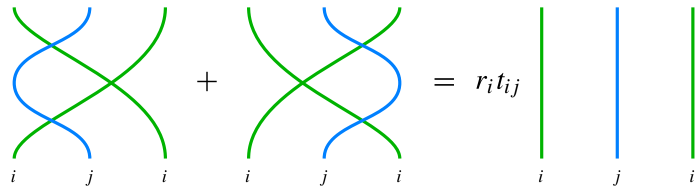
(4) 

(5) 

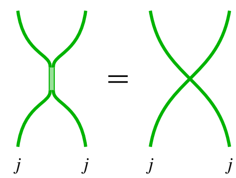
(6) 

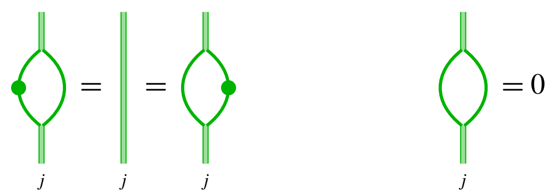
(7)


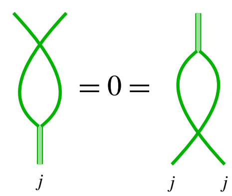
unless $i=k$ and $|i-j|=1$,

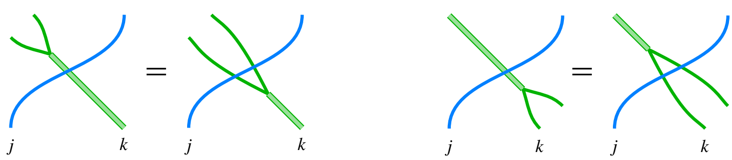
(8)


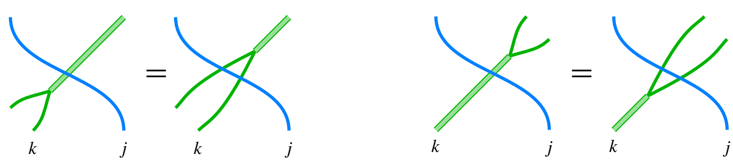
if $|i-j|=1$,

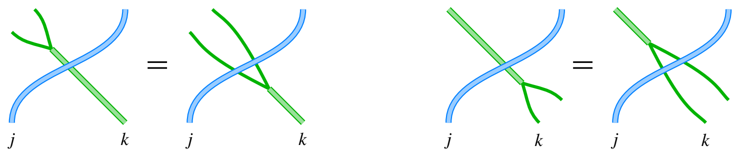
(9)


(10)


(11)


(12)


(13)


(14)


This ends the definition of $\mathfrak{R}(\nu)$.

In Section 2E below we show that $\mathfrak{R}(\nu)$ acts on a supercommutative ring.

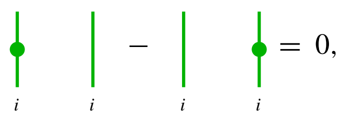
Definition 2.2. We define the monoidal supercategory

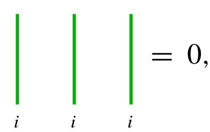
$$\mathfrak{R} = \bigoplus_{\nu \in \mathbb{N}_0[I]} \mathfrak{R}(\nu),$$

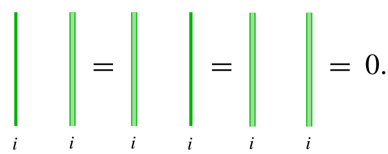
the monoidal structure given by horizontal composition of diagrams.

2B. ***Further relations in*** $\mathfrak{R}(\nu)$**.** We have several consequences of the defining relations.

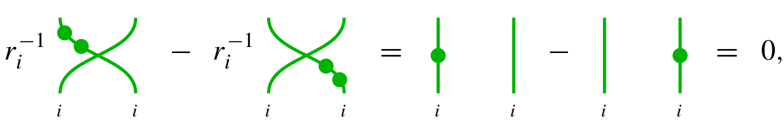
Lemma 2.3. *For all* $i \in I$,

(15)


(16)


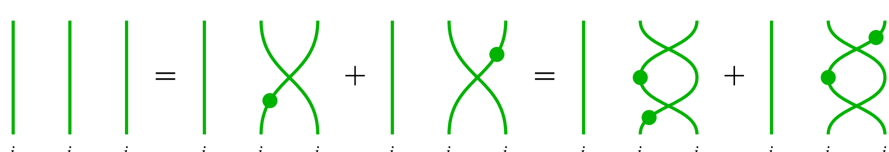
(17)


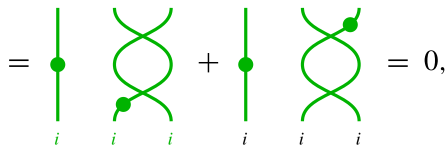
Proof. By (2) and (6),



which proves (15).

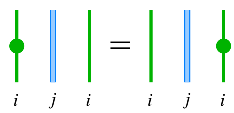
Also,



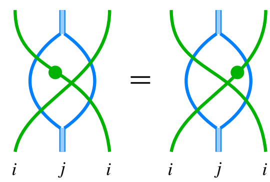


and this proves (16). Relations (17) are an easy consequence of (10) together with (16). □

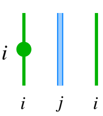
Lemma 2.4. *For all* $i, j \in I$ *with* $|i - j| = 1$,



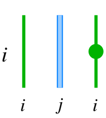
Proof. Start from the equality



Sliding up the dot on the left-hand side using (4) and (1), followed by (8) to pass the ii-crossing to the left, and simplifying using (3) and (10) gives

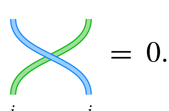
$-r_i t_{ij} t_{ji}$ 

Proceeding similarly on the right-hand side, but sliding the ii-crossing to the right gives

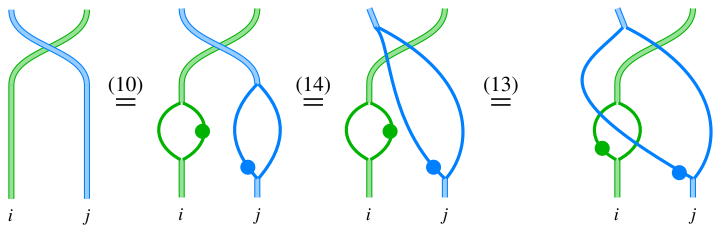
$-r_i t_{ij} t_{ji}$ 

and the claim follows. □

Lemma 2.5. *For all* $i, j \in I$ *with* $|i - j| = 1$,



Proof. We compute:

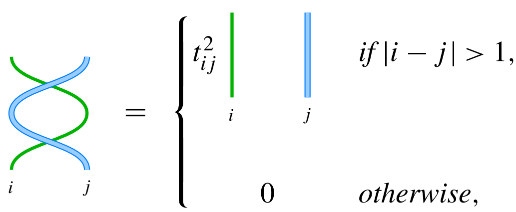


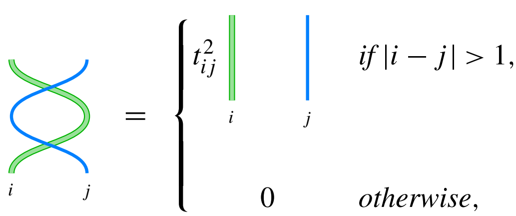
which is zero if $i = j \pm 1$ by (4), (5) and (2). □

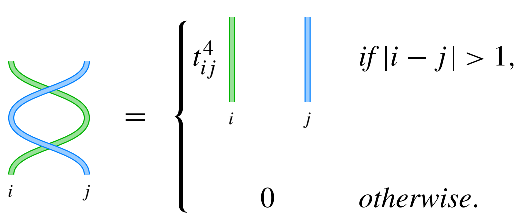
The following are easy consequences of the defining relations of $\mathfrak{R}(\nu)$.

Lemma 2.6. *For all* $i, j \in I$,

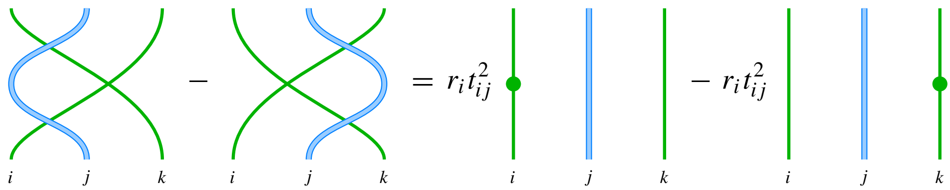
Lemma 2.7. *For all* $i, j \in I$,







Lemma 2.8. *If* $|i-j|=1$ *and* $i=k$,



If $i \neq j \neq k$ *and at least one of the strands is double, then the right hand side is zero.*

Let

$$\mathrm{Seq}(\nu) := \left\{ i_1^{(\varepsilon_1)} \cdots i_r^{(\varepsilon_r)} \in \mathrm{CSeq}(\nu) \mid \varepsilon_s = 1 \right\} \subset \mathrm{CSeq}(\nu).$$

The superalgebra

$$\overline{\mathfrak{R}}(\nu) = \bigoplus_{\boldsymbol{i},\boldsymbol{j} \in \mathrm{Seq}(\nu)} \mathrm{Hom}_{\mathfrak{R}(\nu)}(\boldsymbol{i},\boldsymbol{j}),$$

is the subsuperalgebra of the Hom-superalgebra of $\mathfrak{R}(\nu)$ consisting of all diagrams having only simple strands. If we interpret $\overline{\mathfrak{R}}(\nu)$ as a superalgebra version of a level 2 cyclotomic KLR algebra for $\mathfrak{sl}_n$ then $\mathfrak{R}(\nu)$ can be seen as version of the thick calculus [Khovanov et al. 2012; Stošić 2019] for this superalgebra. It is not hard to see that both the center and the supercenter of $\overline{\mathfrak{R}}(\nu)$ are zero.

2C. *Cyclotomic quotients.* Fix a $\mathfrak{sl}_n$-weight Λ and denote by $R^{\Lambda}(\nu)$, $\overline{\mathfrak{R}}^{\Lambda}(\nu)$ and $\mathfrak{R}^{\Lambda}(\nu)$ the cyclotomic quotients of $R(\nu)$, $\overline{\mathfrak{R}}(\nu)$ and $\mathfrak{R}(\nu)$. The following is immediate.

Lemma 2.9. *If* Λ *is of level* 2 *then the algebras*

$$\overline{\mathfrak{R}}^{\Lambda}(\nu) \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z}) \quad \textit{and} \quad R^{\Lambda}(\nu) \otimes_{\mathbb{Z}} (\mathbb{Z}/2\mathbb{Z})$$

are isomorphic (*after collapsing the* $\mathbb{Z}/2\mathbb{Z}$ *grading of* $\overline{\mathfrak{R}}^{\Lambda}(\nu)$).

We depict a morphism of $\mathfrak{R}^{\lambda}(\nu)$ by decorating the rightmost region of each diagram D with the weight Λ. This defines weights for all regions of D.

The supercategory $\mathfrak{R}^{\Lambda} := \bigoplus_{\nu \in \mathbb{N}_0[I]} \mathfrak{R}^{\Lambda}(\nu)$ is not monoidal anymore, but it is a (left) module category over $\mathfrak{R}$, where $\mathfrak{R}$ acts by adding diagrams of $\mathfrak{R}$ to the left of diagrams from $\mathfrak{R}^{\Lambda}$. This is expressed by a bifunctor

$$\Phi \colon \mathfrak{R} \times \mathfrak{R}^{\lambda} \to \mathfrak{R}^{\lambda}. \tag{18}$$

2D. ***A super 2-category.*** There is a super 2-category around $\mathfrak{R}(\nu)$, paralleling the case of Khovanov–Lauda and Rouquier. An element $\boldsymbol{i} = i_1^{(\varepsilon_1)} \cdots i_r^{(\varepsilon_r)}$ in $\mathrm{CSeq}(\nu)$ corresponds to a root $\alpha_{\boldsymbol{i}} := \sum_s \varepsilon_s \alpha_s$. Let

$$\Lambda(n, d) := \big\{\mu \in \{0, 1, 2\}^n \mid \mu_1 + \cdots + \mu_n = d\big\}.$$

Define $\mathcal{R}(n, d)$ as the super 2-category with objects the elements of $\Lambda(n, d)$ and with morphism supercategories $\mathrm{HOM}_{\mathcal{R}(n,d)}(\mu, \mu')$ the various $\mathfrak{R}(\nu)$. In other words, a 1-morphisms $\mu \to \mu'$ is a sequence $\boldsymbol{i}$ such that $\mu' - \mu = \alpha_{\boldsymbol{i}}$ and the 2-morphism space $\boldsymbol{i} \to \boldsymbol{j}$ is $\mathrm{Hom}_{\mathfrak{R}(\nu)}(\boldsymbol{i}, \boldsymbol{j})$.

Similarly we define the super 2-category $\mathcal{R}^{\Lambda}(n, d)$ by using the cyclotomic quotient with respect with the integral dominant weight Λ. Both super 2-categories $\mathcal{R}^{\Lambda}(n, d)$ have diagrammatic presentations with regions labeled by objects Λ. The 2-morphisms in $\mathcal{R}^{\lambda}(n, d)$ are presented as a collection of 2-morphisms in $\mathcal{R}(n, d)$ with rightmost region decorated with Λ, subjected to the same relations together with the cyclotomic condition. This defines a label for every region of a diagram of $\mathcal{R}^{\Lambda}(n, d)$.

For later use, we denote

$$F_{\boldsymbol{i}}\lambda := F_{i_1^{(\varepsilon_1)} \cdots i_r^{(\varepsilon_r)}}\lambda := F_{i_1}^{(\varepsilon_1)} \cdots F_{i_r}^{(\varepsilon_r)}\lambda$$

the 1-morphisms of $\mathcal{R}^{\Lambda}(n, d)$ and, by abuse of notation, the objects of $\mathfrak{R}^{\Lambda}$.

2E. ***Action on a supercommutative ring.*** We now construct an action of $\mathfrak{R}(\nu)$ on exterior spaces.

2E1. *Demazure operators on an exterior algebra.* Let $V = \bigwedge(y_1, \ldots, y_d)$ be the exterior algebra in d variables. This algebra is naturally graded by word length. Denote by $|z|$ the degree of the homogeneous element z.

The symmetric group $\mathfrak{S}_d$ acts on V by the permutation action,

$$w y_i = y_{w(i)}$$

for all $w \in \mathfrak{S}_d$.

Define operators ∂_i for $i = 1, \ldots, d-1$ on V by the following rules:

$$\partial_i(y_k) = \begin{cases} 1 & i = k, k+1, \\ 0 & \text{otherwise}, \end{cases}$$

$$\text{and} \quad \partial_i(fg) = \partial_i(f)g + (-1)^{|f|} f \partial_i(g),$$

for all $f, g \in V$ such that $fg \neq 0$.

The following can be checked through a simple computation.

Lemma 2.10. *The operators ∂_i satisfy the relations $\partial_i^2 = 0$, $\partial_i\partial_j + \partial_j\partial_i = 0$ if $|i - j| > 1$, and $\partial_i\partial_{i+1}\partial_i = \partial_{i+1}\partial_i\partial_{i+1}$.*

2E2. *An action of $\mathfrak{R}(\nu)$ on supercommutative rings.* For $\boldsymbol{i} \in \mathrm{CSeq}(\nu)$ let

$$P\boldsymbol{i} = \bigwedge(x_{1,1}, x_{1,\varepsilon_1}, \dots, x_{d,1}, x_{d,\varepsilon_d})\boldsymbol{i},$$

be an exterior algebra in $\sum_i \nu_i$ generators, and set

$$P(\nu) = \bigoplus_{\boldsymbol{i} \in \mathrm{CSeq}(\nu)} P\boldsymbol{i}.$$

We extend the action of $\mathfrak{S}_d$ from V to $P(\nu)$ by declaring that

$$w x_{r,1} = x_{w(r),1}, \qquad w x_{r,\varepsilon_r} = x_{w(r),\varepsilon_{r+1}},$$

or $w \in \mathfrak{S}_d$.

Below we denote by $\partial_{u,z}$ the Demazure operator with respect to the variables u and z.

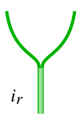
To the object $\boldsymbol{i} \in \mathfrak{R}(\nu)$ we associate the idempotent $\boldsymbol{i} \in P\boldsymbol{i}$. The defining generators of $\mathfrak{R}(\nu)$ act on P as follows. A diagram D acts as zero on $P\boldsymbol{i}$ unless the sequence of labels in the bottom of D is $\boldsymbol{i}$.

- Dots

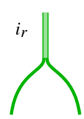
$$: p\boldsymbol{i} \mapsto x_{r,1} p\boldsymbol{i},$$

- Splitters

(19)

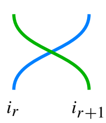


$$: p\boldsymbol{i} \mapsto \partial_{x_{r,1},x_{r,2}}(p)\boldsymbol{i},$$



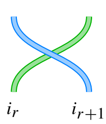
$$: p\boldsymbol{i} \mapsto x_{r,1}\partial_{x_{r,1},x_{r,2}}(p)\boldsymbol{i},$$

- Crossings



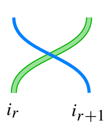
$$: p\boldsymbol{i} \mapsto \begin{cases} r_{i_r}\partial_{x_{r,1},x_{r+1,1}}(p)\boldsymbol{i} & \text{if} \quad i_r = i_{r+1}, \\ (t_{i_{r+1}i_r}x_{r,1} + t_{i_r i_{r+1}}x_{r+1,1})s_r(p\boldsymbol{i}) & \text{if} \quad i_r = i_{r+1}+1, \\ s_r(p\boldsymbol{i}) & \text{else}, \end{cases}$$

(20)

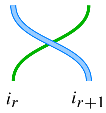


$$: p\boldsymbol{i} \mapsto \begin{cases} 0 & \text{if} \quad i_r = i_{r+1}, \quad \text{or} \quad i_s = i_{s+1}+1, \\ s_r(p\boldsymbol{i}) & \text{else}, \end{cases}$$

(21)



$$: p\boldsymbol{i} \mapsto \begin{cases} 0 & \text{if} \quad i_r = i_{r+1}, \\ f_{2,1}(x_{r,1}, x_{r,2}, x_{r+1,1})s_r(p\boldsymbol{i}) & \text{if} \quad i_s = i_{s+1}+1, \\ s_r(p\boldsymbol{i}) & \text{else}, \end{cases}$$

(22) 

$$: p\boldsymbol{i} \mapsto \begin{cases} 0 & \text{if} \quad i_r = i_{r+1}, \\ f_{1,2}(x_{r,1}, x_{r+1,1}, x_{r+1,2}) s_r(p\boldsymbol{i}) & \text{if} \quad i_s = i_{s+1}+1, \\ s_r(p\boldsymbol{i}) & \text{else,} \end{cases}$$

where

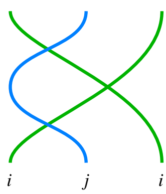
$$f_{2,1}(x_{r,1}, x_{r,2}, x_{r+1,1}) \\ = t_{i_r i_{r+1}} t_{i_{r+1} i_r} x_{r,1} x_{r+1,1} + t_{i_r i_{r+1}} t_{i_{r+1} i_r} x_{r,1} x_{r,2} + t^2_{i_{r+1} i_r} x_{r,2} x_{r+1,1}.$$

$$f_{1,2}(x_{r,1}, x_{r+1,1}, x_{r+1,2}) \\ = -t^2_{i_r i_{r+1}} x_{r,1} x_{r,2} + t_{i_r i_{r+1}} t_{i_{r+1} i_r} x_{r,2} x_{r+1,1} - t_{i_{r+1} i_r} t_{i_{r+1} i_r} x_{r,1} x_{r+1,1}.$$

Proposition 2.11. *The assignment above defines an action of* $\mathfrak{R}(\nu)$ *on* $P(\nu)$.

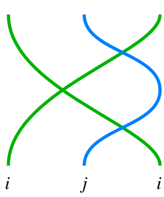
Proof. By a long and rather tedious computation one can check that the operators above satisfy the defining relations of $\mathfrak{R}(\nu)$.

The relations involving the action of the generators of $\overline{\mathfrak{R}}(\nu)$ are easy to check by direct computation. For example, for $\nu = 2i + j$, with $j = i + 1$ we have



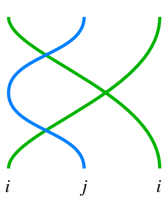
$$(f) = (t_{ij}x_1 + t_{ji}x_2) s_1 r_i \partial_2 s_1(f),$$

and

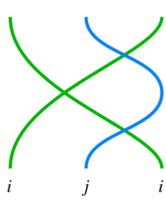


$$(f) = s_2 r_i \partial_1 (t_{ij}x_2 + t_{ji}x_3) s_2(f) = r_i t_{ij} f - (t_{ij}x_1 + t_{ji}x_2) s_1 r_i \partial_2 s_1(f),$$

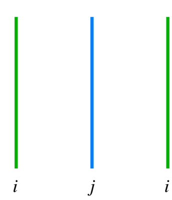
and so, for any $f(x_1, x_2, x_3) \in Piji$,



$(f) \ +$

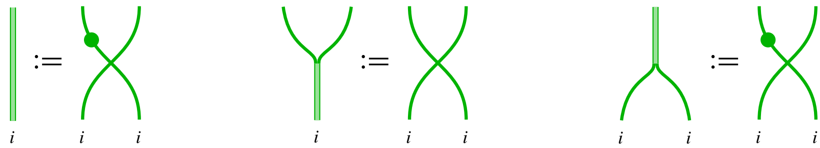


$(f) \ = \ r_i t_{ij}$

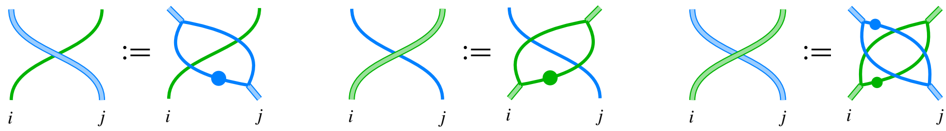


$(f).$

Setting as in [Khovanov et al. 2012],



and



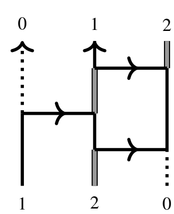
then it follows that the action of the generators of $\mathfrak{R}(\nu)$ on $P(\nu)$ is given by the operators (19), (20), (21) and (22) and satisfy the defining relations of $\mathfrak{R}(\nu)$. □

3. A topological invariant

In [Tubbenhauer 2014] q-skew Howe duality is used to show how to write as a web in a form that uses only the lower part of $U_q(\mathfrak{gl}_k)$. In this language, the formula for the $\mathfrak{sl}_2$-commutator becomes one of Lusztig's higher quantum Serre relations [1993, §7]. It is also proved in [Tubbenhauer 2014] that this results in a well defined evaluation of closed webs allowing to write any link diagram as a linear combination of words in the various F_i in $U^- := U_q^-(\mathfrak{gl}_k)$.

This allows a categorification of webs using only (cyclotomic) KLR algebras [Khovanov and Lauda 2009; Rouquier 2008] instead of the whole 2-quantum group $\mathcal{U}(\mathfrak{gl}_k)$ [Khovanov and Lauda 2010; Rouquier 2008]. In this context, the unit and counit maps of the several adjunctions in $\mathcal{U}(\mathfrak{gl}_k)$ that are used as differentials in the Khovanov–Rozansky chain complex can be written as composition with elements of the KLR algebra. Taking cyclotomic KLR algebras of level 2 gives Khovanov homology. The approach in [Tubbenhauer 2014] is easily adapted to tangles, which we do in this section for level 2 in the context of the supercategories introduced in Section 2.

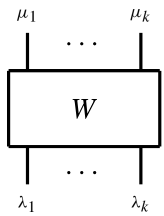
3A. ***Supercategorification of $\mathfrak{gl}_2$-webs and flat tangles.*** Our webs have strands labeled from {0, 1, 2} which we depict as "invisible," "simple," and "double," as in the example below. All the strands point either up or to the right and sometimes we omit the orientations in the pictures.



For $\lambda = (\lambda_1, \dots, \lambda_k) \in \{0, 1, 2\}^k$ and $\epsilon \in \{0, 1\}$ with $|\lambda| = 2\ell + \epsilon$, we put $\Lambda = (2)^\ell \epsilon = (2, \dots, 2, \epsilon, 0, \dots, 0)$ and we define

$$\mathfrak{W}(\lambda) = \mathrm{HOM}_{\mathcal{R}^\Lambda(k,|\lambda|)}(\Lambda, \lambda).$$

Let W be a $\mathfrak{gl}_2$-web with all ladders pointing to the right. Suppose that W has the bottom boundary labeled λ and the top boundary labeled μ, with $\lambda, \mu \in \{0, 1, 2\}^k$ and $|\lambda| = |\mu|$. We write W as a word in the F_i in $U_q^-(\mathfrak{gl}_k)$ applied to a vector v_λ of $\mathfrak{gl}_k$-weight λ.



$$= F_{i_1} \cdots F_{i_r}(v_\lambda).$$

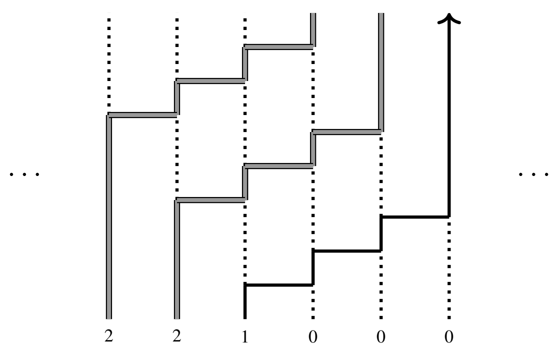
This gives a 1-morphism $F(W)$ in $\mathcal{R}(k, |\lambda|)$. Composition of 1-morphisms in $\mathcal{R}(k, |\lambda|)$ defines a superfunctor

$$\mathfrak{F}(W)\colon \mathfrak{W}(\lambda) \to \mathfrak{W}(\mu).$$

If λ is dominant and μ is antidominant then $\mathfrak{F}(W)$ is a superfunctor from $\Bbbk$-smod to $\Bbbk$-smod that is, a direct sum of grading shifts of the identity superfunctor. In this case, there is a canonical 1-morphism $F_{\mathrm{can}}(W)$ in $\mathrm{Hom}_{\mathcal{R}^\Lambda(k,|\lambda|)}(\lambda, \mu)$

(23) $$F_{\mathrm{can}} = F_{(k-\ell-1)^{(2)}\dots(1)^{(2)}} \cdots F_{(k-3)^{(2)}\dots(\ell-1)^{(2)}} F_{(k-2)^{(2)}\dots\ell^{(2)}} F_{(k-1)^{(\epsilon)}\dots(\ell+1)^{(\epsilon)}} (2)^\ell \epsilon,$$

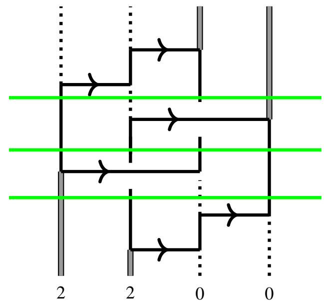
which in terms of webs takes the form of the following example:



We have that $\mathfrak{F}(W) = \mathrm{Hom}_{\mathcal{R}^\lambda(k,|\lambda|)}(\lambda, \mu)$ is isomorphic to the graded $\Bbbk$-supervector space $\mathrm{Hom}_{\mathfrak{R}^\Lambda}(F_{\mathrm{can}}(W), F(W))$.

3B. ***The chain complex.*** As explained in [Tubbenhauer 2014] any oriented tangle diagram T can be written in the form of a web W_T with all horizontal strands pointing to the right. In this case we say that T *is in F-form*.

Example 3.1. For the Hopf link we have the following web diagram.



Suppose the bottom boundary of W_T is $(\lambda_1, \dots, \lambda_k)$ and the top boundary is $(\mu_1, \dots, \mu_k)$. Let $\mathrm{Kom}(\lambda, \mu)$ be the category of complexes of

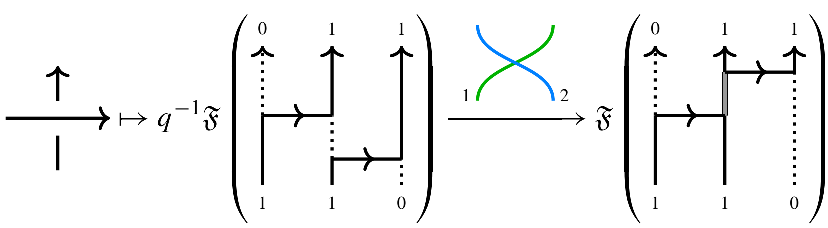
$$\mathrm{HOM}_{\mathcal{R}(k,|\lambda|)}(\mathfrak{W}(\lambda), \mathfrak{W}(\mu))$$

generated monoidally by tensor products of complexes of length 2, and $\mathrm{Kom}_{/h}(\lambda, \mu)$ its homotopy category (these are not supercategories). The usual constructions with chain complexes (homomorphisms, homotopies, cones, etc.) work in the same way as with nonsupercategories. Since we are in a supercategory, some signs have to be introduced (further details will appear in a follow-up paper). To each tangle in F-form as above we associate an object in $\mathrm{Kom}_{/h}(\lambda, \mu)$ as follows.

We first chop the diagram vertically in such way that each slice contains either a web without crossings, or a single crossing together with vertical pieces (as in Example 3.1). Each slice then gives either a superfunctor or a complex of superfunctors, as explained below. By composition we get a complex $\mathfrak{F}(W_T)$ of superfunctors from $\mathfrak{W}(\lambda)$ to $\mathfrak{W}(\mu)$.

3B1. *Basic tangles.*

- If T is a flat tangle, then we're done by Section 3A.
- To the positive crossing we associate the chain complex

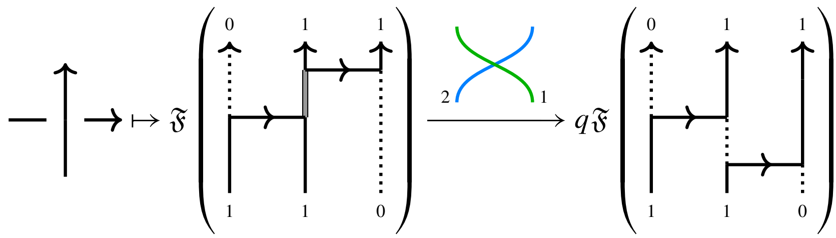
(24) 

with the leftmost term in homological degree zero. Algebraically this can be written

$$\beta_+ \mapsto q^{-1} F_1 F_2(1, 1, 0) \xrightarrow{\tau_1} F_2 F_1(1, 1, 0),$$

where τ is the diagram above.

- To the negative crossing we associate the chain complex

(25) 

with the rightmost term in homological degree zero. Algebraically

$$\beta_- \mapsto F_2F_1(1,1,0) \xrightarrow{\tau_1} qF_1F_2(1,1,0).$$

Remark 3.2. Caution should be taken when applying (24) and (25): when passing from a tangle diagram to it's F-form some crossings may change from positive to negative and vice versa. To have an invariant of all tangles some grading shifts have to be introduced locally whenever this occurs. We shift (25) by -1 in the q-grading and 1 in the homological grading when it comes from a positive crossing and the opposite whenever (25) comes from a positive crossing.

3B2. *The normalized complex.* Let $n_\pm$ be the number of positive/negative crossings in W_T and let $w = n_+ - n_-$ be the writhe of W_T. We define the normalized complex

$$\mathfrak{F}(W_T) := q^{2w}\overline{\mathfrak{F}}(W_T). \tag{26}$$

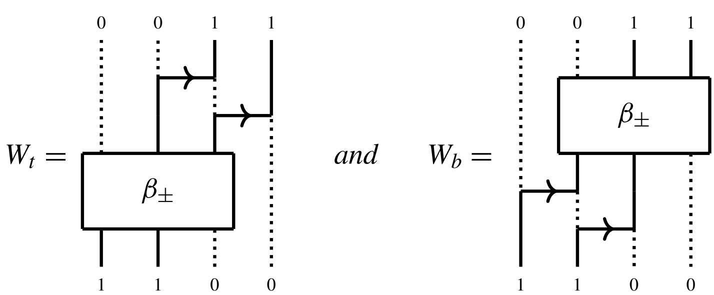
3C. ***Topological invariance.***

Theorem 3.3. *For every tangle diagram T the homotopy type of $\mathfrak{F}(W_T)$ is invariant under the Reidemeister moves.*

Theorem 3.4. *For every link L the homology of $\mathfrak{F}(L)$ is a $\mathbb{Z}$-graded supermodule over $\mathbb{Z}$ whose graded Euler characteristic equals the Jones polynomial.*

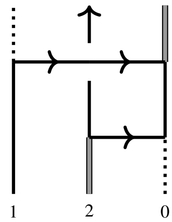
Proof of Theorem 3.3. The following is immediate.

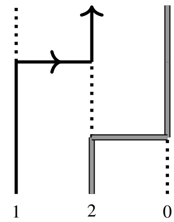
Lemma 3.5. *For $\beta_\pm$ a positive/negative crossing let W_t and W_b be the following tangles in F-form:*



Then the complexes $\mathfrak{F}(W_t)$ *and* $\mathfrak{F}(W_b)$ *are isomorphic.*

Lemma 3.6 (Reidemeister I). *Consider diagrams* D_1^+ *and* D_0 *that differ as below.*

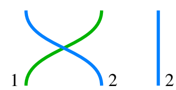
$D_1^+ =$ 

$D_0 =$ 

Then $\mathfrak{F}(D_1^+)$ *and* $\mathfrak{F}(D_0)$ *are isomorphic in* $\mathrm{Kom}_{/h}\big((1,2,0),(0,1,2)\big)$.

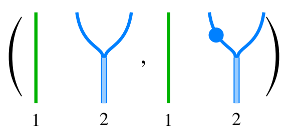
Proof. We have

$$\overline{\mathfrak{F}}(D_1^+) = q^{-1}F_1F_2F_2(1,2,0) \longrightarrow F_1F_1F_2(1,2,0).$$



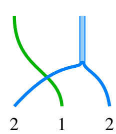
The first term is isomorphic to $F_1F_2^{(2)}(1,2,0) \oplus q^{-2}F_1F_2^{(2)}(1,2,0)$ via the map

$$F_1F_2^{(2)}(1,2,0)\oplus q^{-2}F_1F_2^{(2)}(1,2,0) \xrightarrow[\simeq]{} q^{-1}F_1F_2^2(1,2,0),$$



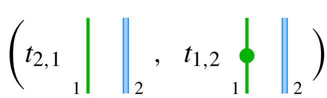
while for the second term there is an isomorphism

$$F_2F_1F_2(1,2,0) \xrightarrow[\simeq]{} F_1F_2^{(2)}(1,2,0),$$



so that $\overline{\mathfrak{F}}(D_1^+)$ is isomorphic to the complex

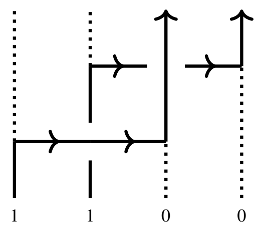
$$\begin{pmatrix} F_1F_2^{(2)}(1,2,0) \\ q^{-2}F_1F_2^{(2)}(1,2,0) \end{pmatrix} \longrightarrow F_1F_2^{(2)}(1,2,0).$$

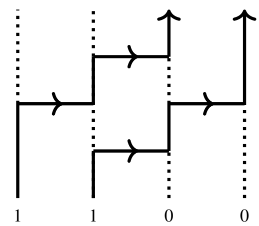


By Gaussian elimination one gets that the complex $\overline{\mathfrak{F}}(D_1^+)$ is homotopy equivalent to the one term complex $q^{-2}F_1F_2^{(2)}(1,2,0)$ concentrated in homological degree zero, which after normalization is $\mathfrak{F}(D_0)$. □

The other types of Reidemeister I move can be verified similarly. For example, replacing the positive crossing by a negative crossing in Lemma 3.6 and using the inverses of the various isomorphisms above results in a complex isomorphic to $\overline{\mathfrak{F}}(D_1^-)$ that is homotopy equivalent to the 1-term complex $q^2F_1F_2^{(2)}(1,2,0)$ concentrated in homological degree zero.

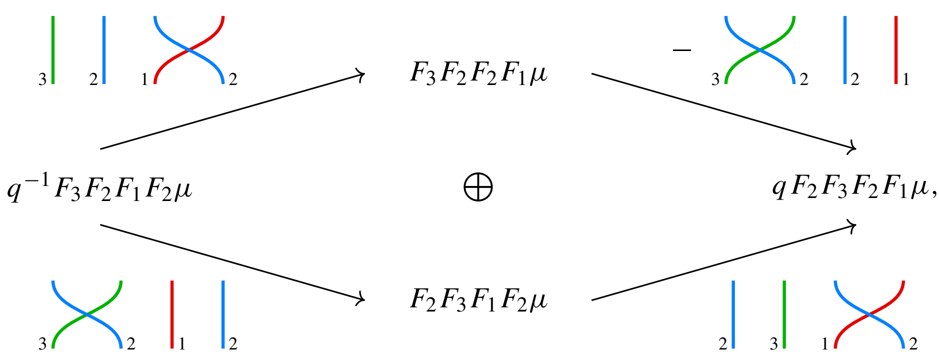
Lemma 3.7 (Reidemeister IIa). *Consider diagrams D_1 and D_0 that differ as below.*

$D_1 =$ 

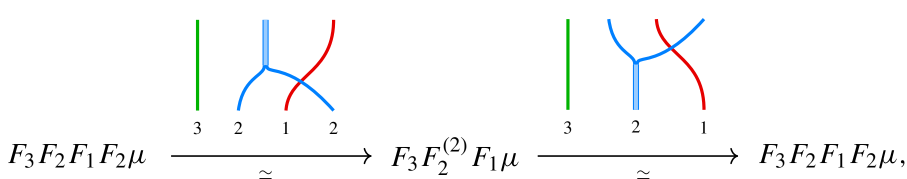
$D_0 =$ 

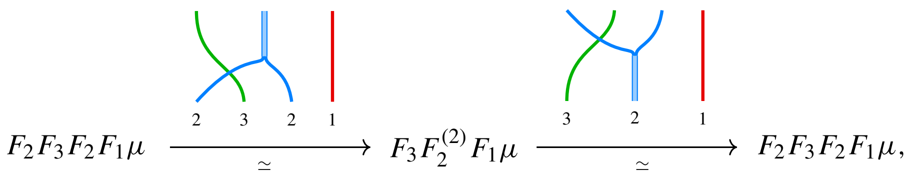
Then $\mathfrak{F}(D_1)$ and $\mathfrak{F}(D_0)$ are isomorphic in $\mathrm{Kom}_{/h}\big((1,1,0,0),(0,0,1,1)\big)$.

Proof. In the following we write μ instead of $(1,1,0,0)$. The complex $\overline{\mathfrak{F}}(D_1)$ is

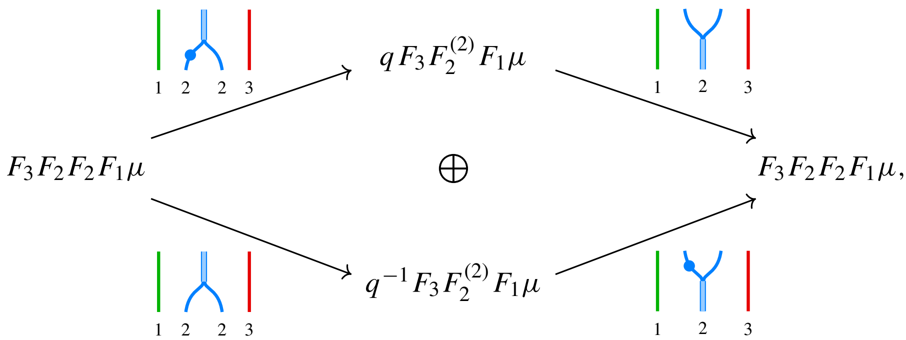


From the isomorphisms

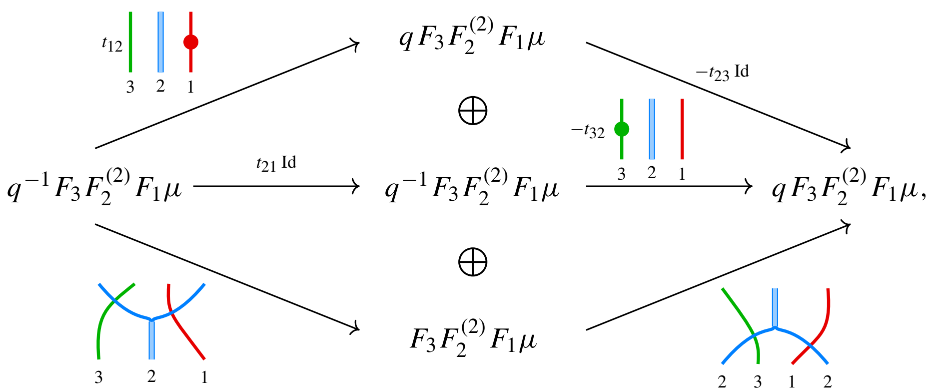




and



and simplifying the maps using the relations in $\mathfrak{R}(\nu)$ one gets that $\mathfrak{F}(D_1)$ is isomorphic to the complex



By Gaussian elimination of the acyclic two-term complexes

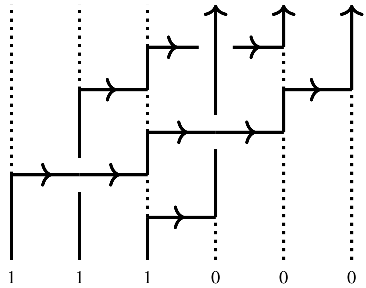
$$q^{-1}F_3F_2^{(2)}F_1\mu \xrightarrow{t_{21}\,\mathrm{Id}} q^{-1}F_3F_2^{(2)}F_1\mu \quad \text{and} \quad qF_3F_2^{(2)}F_1\mu \xrightarrow{-t_{23}\,\mathrm{Id}} qF_3F_2^{(2)}F_1\mu$$

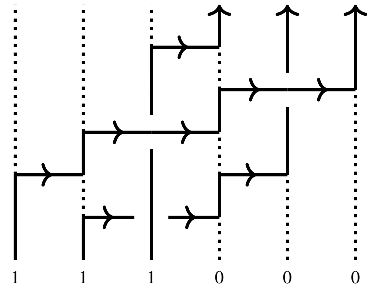
one obtains that $\overline{\mathfrak{F}}(D_1)$ is homotopy equivalent to the complex

$$0 \longrightarrow F_3F_2^{(2)}F_1\mu \longrightarrow 0,$$

with the middle-term in homological degree zero. □

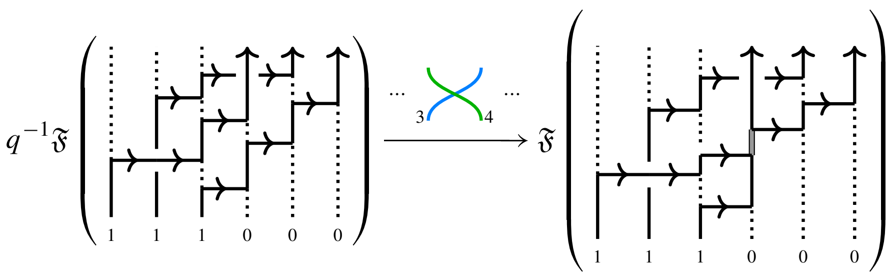
Lemma 3.8 (Reidemeister III). *Consider diagrams D_L and D_R that differ as below.*

$D_L =$ 

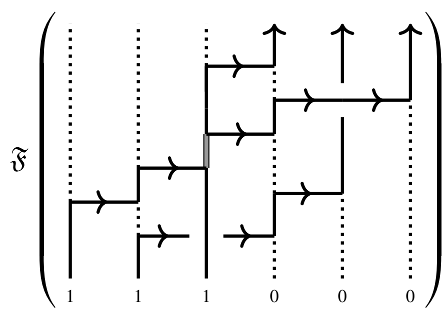
$D_R =$ 

Then $\mathfrak{F}(D_L)$ and $\mathfrak{F}(D_R)$ are isomorphic in $\mathrm{Kom}_{/h}\big((1,1,1,0,0,0),(0,0,0,1,1,1)\big)$.

Proof. The proof is inspired by [Putyra 2014a, Lemma 7.9] (see also [Putyra 2014b, §4.3.3] for further details). The complex associated to D_L is the mapping cone of the map



An easy exercise shows that the second complex is isomorphic to the complex



In [Putyra 2014b, §4.3.3] it is explained in detail how to use an isomorphism like this together with the maps associated to two Reidemeister 2 moves on the first complex to prove that $\mathfrak{F}(D_L)$ is homotopy equivalent to $\mathfrak{F}(D_R)$. □

This finishes the proof of Theorem 3.3. □

3D. ***Not even Khovanov homology.*** We now show that for links the invariant $\mathcal{H}(L)$ is distinct from even Khovanov homology and shares common properties with odd Khovanov homology.

3D1. *Reduced homology.*

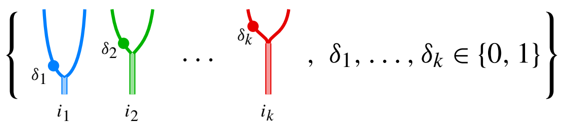
Theorem 3.9. *For every link L there is an invariant $H_{\text{reduced}}(L)$ with the property*

$$H(L) \simeq q\,H_{\text{reduced}}(L) \oplus q^{-1} H_{\text{reduced}}(L).$$

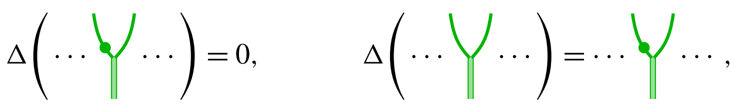
The proof of Theorem 3.9 follows a reasoning analogous to the proof of Theorem 3.2.A. in [Shumakovitch 2014], for the analogous decomposition for Khovanov homology over $\mathbb{Z}/2\mathbb{Z}$ in terms of reduced Khovanov homology.

Before proving the theorem we do some preparation. Recall that for D a diagram of L the chain groups of $\mathfrak{F}(D)$ are the various $\Bbbk$-supervector spaces $\mathrm{Hom}_{\mathcal{R}^\wedge}(F_{\text{can}}, F(W))$, where W runs over all the resolutions of D.

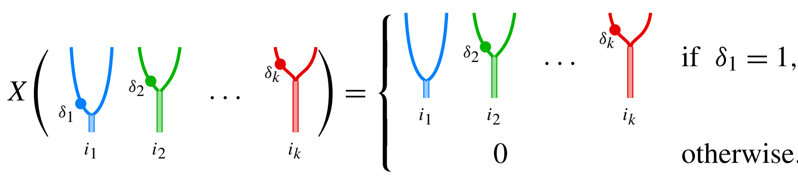
If we write $F_{\text{can}} = F_{i_1^{(2)} i_2^{(2)} \cdots i_k^{(2)}}$ then $\mathrm{Hom}_{\mathcal{R}^\wedge}(F_{\text{can}}, F_{i_1 i_1 i_2 i_2 \cdots i_k i_k})$ is spanned by



Introduce linear maps X and Δ on $\mathrm{Hom}_{\mathcal{R}^\wedge}(F_{\text{can}}, F_{i_1 i_1 i_2 i_2 \cdots i_k i_k})$ as follows. Map Δ is defined on the factors as



and extended to $\mathrm{Hom}_{\mathcal{R}^\wedge}(F_{\text{can}}, F_{i_1 i_1 i_2 i_2 \cdots i_k i_k})$ using the Leibniz rule. The map X is defined by



Since

$$\mathrm{Hom}_{\mathcal{R}^\wedge}(F_{\text{can}}, F(W)) \simeq \mathrm{Hom}_{\mathcal{R}^\wedge}(F_{\text{can}}, F_{i_1 i_1 i_2 i_2 \cdots i_k i_k}) \times \mathrm{Hom}_{\mathcal{R}^\wedge}(F_{i_1 i_1 i_2 i_2 \cdots i_k i_k}, F(W))$$

the maps Δ and X induce maps on $\mathrm{Hom}_{\mathcal{R}^\wedge}(F_{\text{can}}, F(W))$, denoted by the same symbols.

Lemma 3.10. *Both maps X and Δ commute with the differential of $\mathfrak{F}(D)$, $\Delta^2 = 0$, and moreover $X\Delta + \Delta X = \mathrm{Id}_{\mathfrak{F}(D)}$.*

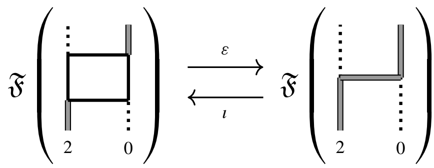
Proof. Straightforward. □

Proof of Theorem 3.9. We have that Δ is acyclic and therefore

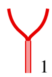
$$\mathfrak{F}(D) \simeq \ker(\Delta) \oplus q^2 \ker(\Delta),$$

and so the claim follows by setting $\mathfrak{F}_{\text{reduced}}(D) = q\ker(\Delta)$. □

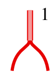
3D2. *A chronological Frobenius algebra.* We now examine the behavior of the functor $\mathfrak{F}$ under merge and splitting of circles. First define maps ι and ε,



as



$$\iota\colon F_1^{(2)}(2,0) \longrightarrow F_1^2(2,0)$$

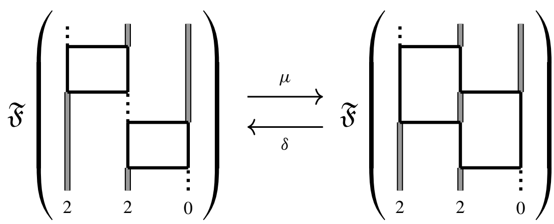


$$\varepsilon\colon F_1^2(2,0) \longrightarrow F_1^{(2)}(2,0).$$

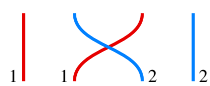
Note that, contrary to [Ozsváth et al. 2013], $p(\iota) = 1$ and $p(\varepsilon) = 0$.

We now consider the following two cases (a) and (b) below.

(a)

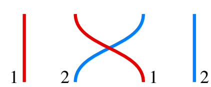


The maps μ and δ are given by



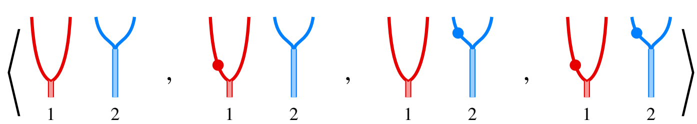
$$\mu\colon F_1^2F_2^2(2,2,0) \longrightarrow F_1F_2F_1F_2(2,2,0),$$

and



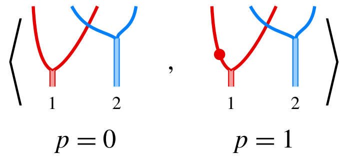
$$\delta\colon F_1F_2F_1F_2(2,2,0) \longrightarrow F_1^2F_2^2(2,2,0).$$

We have $p(\mu) = 0$ and $p(\delta) = 1$. Decomposing $F_1^2F_2^2(2,2,0)$ and $F_1F_2F_1F_2(2,2,0)$ into a direct sum of several copies of $F_1^{(2)}F_2^{(2)}(2,2,0)$ with the appropriate grading shifts we fix bases

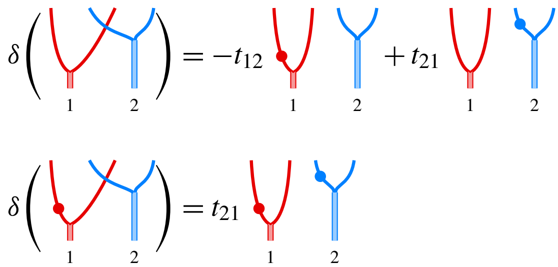


$p = 0$ $\quad$ $p = 1$ $\quad$ $p = 1$ $\quad$ $p = 0$

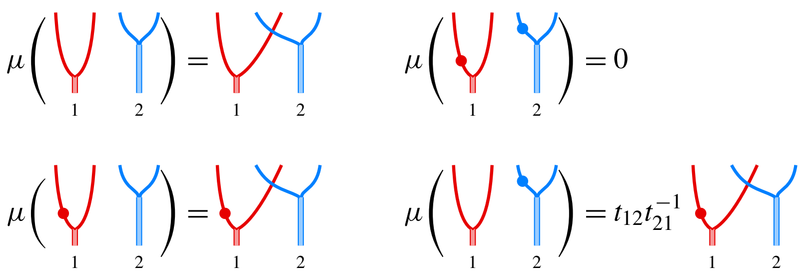
of $F_1^2 F_2^2(2, 2, 0)$, and



of $F_1 F_2 F_1 F_2(2, 2, 0)$. Then we compute



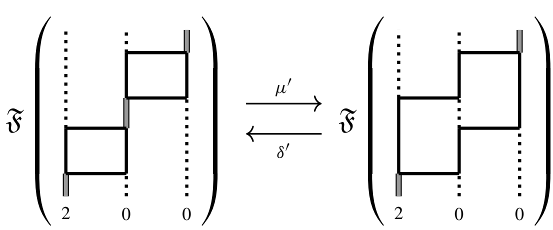
and



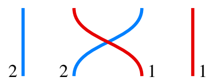
Using this one sees that easily that $\mu\delta = 0$, as in the case of the odd Khovanov homology of [Ozsváth et al. 2013].

Setting to 1 all t_{ij} and renaming $\langle 1, a_1, a_2, a_1 \wedge a_2 \rangle$ the basis vectors of $F_1^2 F_2^2(2,0,0)$ and $\langle 1, a_1 = a_2 \rangle$ the basis vectors of $F_1 F_2 F_1 F_2(2, 0, 0)$ one can give the maps δ, μ, $\imath$ and ε a form that coincides with the corresponding maps in [Ozsváth et al. 2013, §1.1]. Note though, that while the parities of δ and μ coincide with the corresponding maps in [Ozsváth et al. 2013], the parities of $\imath$ and ε are reversed with respect to [Ozsváth et al. 2013].

(*b*)

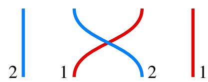


The maps μ' and δ' are given by



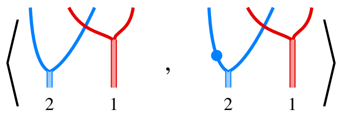
$$\mu' \colon F_2^2F_1^2(2,0,0) \longrightarrow F_2F_1F_2F_1(2,0,0),$$

and



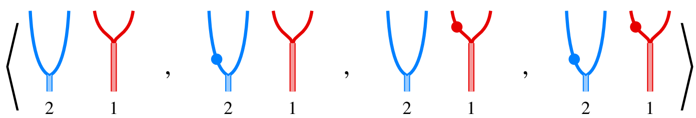
$$\delta' \colon F_2F_1F_2F_1(2,0,0) \longrightarrow F_2^2F_1^2(2,0,0).$$

Proceeding as above we fix a basis



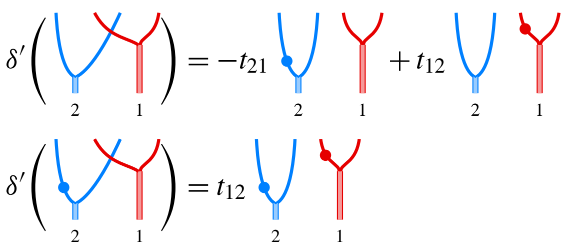
$p=1$ $\qquad$ $p=0$

of $F_2F_1F_2F_1(2,0,0)$ and

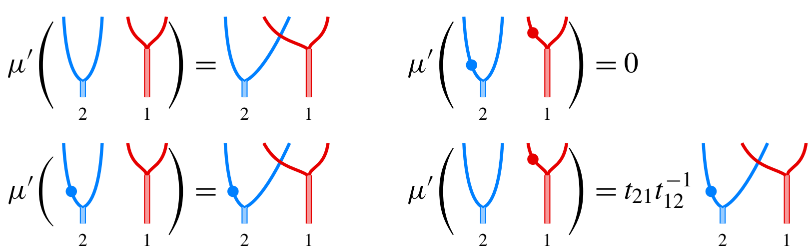


$p=0$ $\qquad$ $p=1$ $\qquad$ $p=1$ $\qquad$ $p=0$

of $F_2^2F_1^2(2,2,0)$, to get



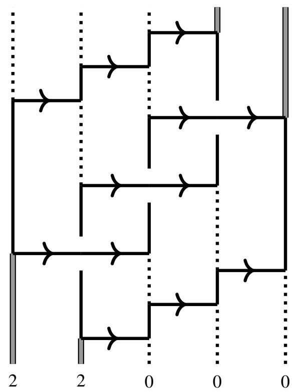
and



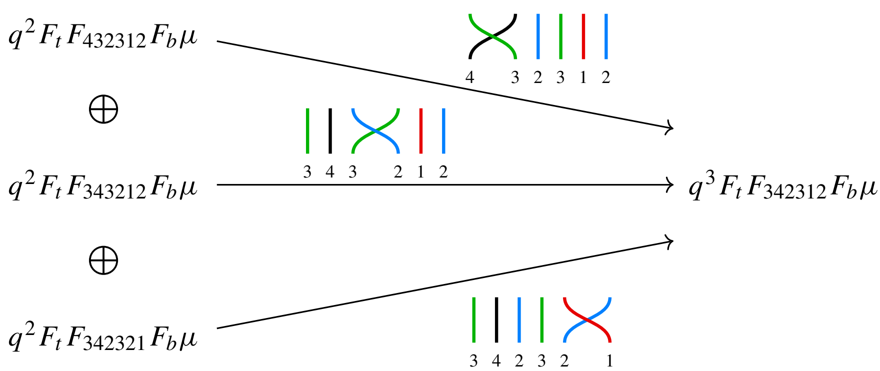
In this case we also have $\mu'\delta' = 0$.

Contrary to the previous case, we have $p(\mu') = 1$ and $p(\delta') = 0$. The maps μ' and δ' can also be made to agree with [Ozsváth et al. 2013], but the parity is reversed (as with ι and ε above).

3D3. *A sample computation.* We now compute the homology of the left-handed trefoil T in its lowest and highest homological degrees. Consider the following presentation of T,



The computation of $H_0(T)$ is fairly simple: up to an overall degree shift it is the homology in degree 1 of the complex

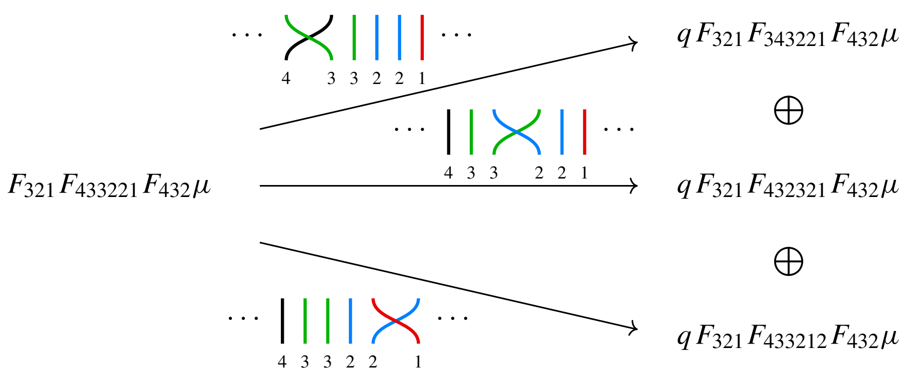
(27)


The three terms in homological degree zero are isomorphic to $F_{43^{(2)}2^{(2)}1}$. Composing the isomorphisms from $F_{43^{(2)}2^{(2)}1}$ to F_{432312}, F_{343212} and to F_{342321} with the corresponding maps above gives three maps that differ by a sign.

By inspection, one sees that up to a sign, these three maps are equal to the map δ from the case (a) in the previous subsection. The cokernel map in (27) is therefore two-dimensional. Adding the degree shifts one obtains

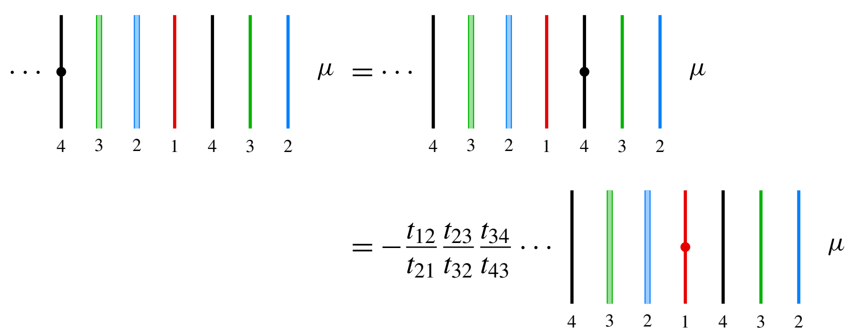
$$H_0(T) = q^{-1}\Bbbk \oplus q^{-3}\Bbbk.$$

We now compute $H_{-3}(H)$. Up to an overall degree shift it is computed as the homology in degree zero of the complex



Here $\mu = (2, 2, 0, 0, 0)$ and the factors F_{321} and F_{432} are the upper and lower closures of the diagram. We write F_t for F_{321} and F_b for F_{432} and sometimes we write $F_tF_{433221}F_b\mu$ instead of $F_{321}F_{433221}F_{432}\mu$, etc., and we only depict the pertinent part of the morphisms.

In the following we will use the identities

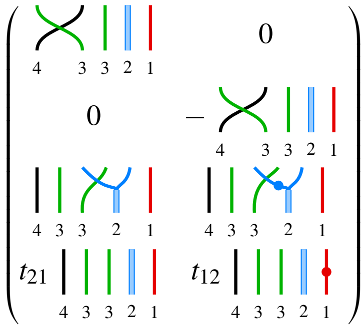
(28)


The first equality follows from Lemma 2.4 after using (3) on the second strand labeled 4 to pull it to the left. The second equality can be checked by a applying (3) three times.

Coming back to $H_{-3}(T)$ we apply the isomorphisms

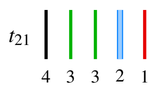
$$F_{433221} \simeq qF_{4332^{(2)}1} \oplus q^{-1}F_{4332^{(2)}1},$$
$$F_{343221} \simeq qF_{3432^{(2)}1} \oplus q^{-1}F_{3432^{(2)}1},$$
$$F_{433212} \simeq F_{4332^{(2)}1},$$

to obtain the isomorphic complex



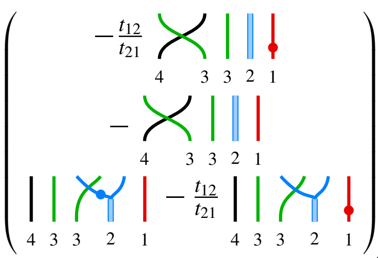
$$\begin{pmatrix} q F_t F_{4332^{(2)}1} F_b \mu \\ q^{-1} F_t F_{4332^{(2)}1} F_b \mu \end{pmatrix} \longrightarrow \begin{pmatrix} q^2 F_t F_{3432^{(2)}1} F_b \mu \\ F_t F_{3432^{(2)}1} F_b \mu \\ q F_t F_{432321} F_b \mu \\ q F_t F_{4332^{(2)}1} F_b \mu \end{pmatrix}.$$

By Gaussian elimination of the acyclic complex



$$q F_t F_{4332^{(2)}1} F_b \mu \longrightarrow q F_t F_{4332^{(2)}1} F_b \mu.$$

we obtain the homotopy equivalent complex

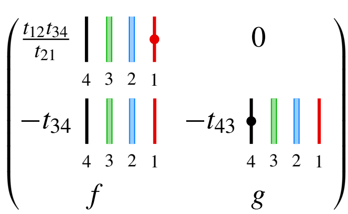


$$q^{-1} F_t F_{4332^{(2)}1} F_b \mu \longrightarrow \begin{pmatrix} q^2 F_t F_{3432^{(2)}1} F_b \mu \\ F_t F_{3432^{(2)}1} F_b \mu \\ q F_t F_{432321} F_b \mu \end{pmatrix}.$$

Applying the isomorphisms

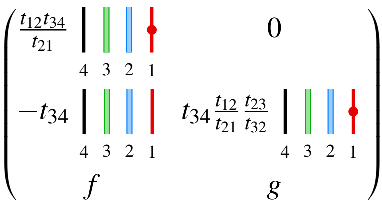
$$F_{4332^{(2)}1} \simeq q F_{43^{(2)}2^{(2)}1} \oplus q^{-1} F_{43^{(2)}2^{(2)}1} \tag{29}$$

and $F_{3432^{(2)}1} \simeq F_{43^{(2)}2^{(2)}1}$ gives the isomorphic complex



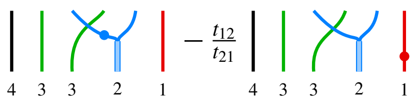
$$\begin{pmatrix} F_t F_{43^{(2)}2^{(2)}1} F_b \mu \\ q^{-2} F_t F_{43^{(2)}2^{(2)}1} F_b \mu \end{pmatrix} \longrightarrow \begin{pmatrix} q^2 F_t F_{43^{(2)}2^{(2)}1} F_b \mu \\ F_t F_{43^{(2)}2^{(2)}1} F_b \mu \\ q F_t F_{432321} F_b \mu \end{pmatrix},$$

or

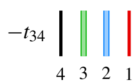


$$\begin{pmatrix} F_tF_{43^{(2)}2^{(2)}1}F_b\mu \\ q^{-2}F_tF_{43^{(2)}2^{(2)}1}F_b\mu \end{pmatrix} \longrightarrow \begin{pmatrix} q^2F_tF_{43^{(2)}2^{(2)}1}F_b\mu \\ F_tF_{43^{(2)}2^{(2)}1}F_b\mu \\ qF_tF_{432321}F_b\mu \end{pmatrix},$$

where f (resp. g) is the composite of the map from $F_{43^{(2)}2^{(2)}1}$ (resp. $q^{-2}F_{43^{(2)}2^{(2)}1}$) to $q^{-1}F_{4332^{(2)}1}$ in (29) and



Gaussian elimination of the acyclic complex

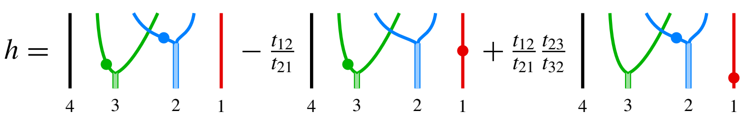


$$F_tF_{43^{(2)}2^{(2)}1}F_b\mu \longrightarrow F_tF_{43^{(2)}2^{(2)}1}F_b\mu,$$

yields the homotopy equivalent complex

$$q^{-2}F_tF_{43^{(2)}2^{(2)}1}F_b\mu \xrightarrow{\begin{pmatrix} 0 \\ h \end{pmatrix}} \begin{pmatrix} q^2F_tF_{43^{(2)}2^{(2)}1}F_b\mu \\ qF_tF_{432321}F_b\mu \end{pmatrix},$$

where



Since we are only interested in the lowest homological degree we restrict to considering the complex

$$q^{-2}F_tF_{43^{(2)}2^{(2)}1}F_b\mu \xrightarrow{h} qF_tF_{432321}F_b\mu.$$

Finally, applying the isomorphism $F_tF_{432321}F_b \simeq F_tF_{4332^{(2)}1}F_b$ results in the isomorphic complex

$$q^{-2}F_tF_{43^{(2)}2^{(2)}1}F_b\mu \xrightarrow{0} qF_tF_{432321}F_b\mu.$$

Adding the shift corresponding to the normalization (26), and using the fact that $F_t F_{43^{(2)}2^{(2)}1} F_b \mu$ is a $\Bbbk$-supervector space of graded dimension $q + q^{-1}$, yields

$$H_{-3}(T) = q^{-7}\Bbbk \oplus q^{-9}\Bbbk,$$

which agrees with the odd Khovanov homology of T.

4. Further properties of $\mathfrak{R}$

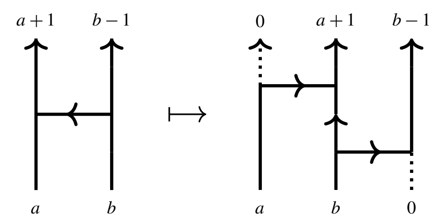
In this section we sketch several of its higher representation theory properties of $\mathfrak{R}$, some of them we have used in the previous section.

4A. ***Supercategorical action on $\mathcal{R}^\Lambda(k, d)$.*** Given a $\mathfrak{gl}_n$-weight $\Lambda = (\Lambda_1, \ldots, \Lambda_n)$ we write $\overline{\Lambda} = (\Lambda_1 - \Lambda_2, \ldots, \Lambda_{n-1} - \Lambda_n)$ for the corresponding $\mathfrak{sl}_n$-weight. The superalgebra $\mathfrak{R}^\Lambda(\nu)$ for $\mathfrak{gl}_k$ is defined to be the same as the superalgebra $\mathfrak{R}^{\overline{\Lambda}}(\nu)$ for $\mathfrak{sl}_k$.

We now explain how the bifunctor $\Phi\colon \mathfrak{R} \times \mathfrak{R}^\Lambda \to \mathfrak{R}^\Lambda$ in (18). gives rise to an action of $\mathfrak{gl}_k$ on $\mathcal{R}^\Lambda(k, d)$ for Λ a dominant integrable $\mathfrak{gl}_k$-weight of level 2 with $\Lambda_1 + \cdots \Lambda_n = d$. A diagram D in $\mathcal{R}^\Lambda(k, d)$ with leftmost region labeled μ defines a web W_D with bottom boundary labeled Λ and with top boundary labeled μ. We denote $f_i, e_i \in U_q(\mathfrak{gl}_k)$ the Chevalley generators.

Behind Tubbenhauer's construction in [Tubbenhauer 2014] there is the observation that the transformation

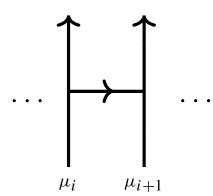
(30)



turns any web into a web with all horizontal edges pointing to the right. This goes through the obvious embedding of $\mathfrak{gl}_k$ into $\mathfrak{gl}_{k+1}$.

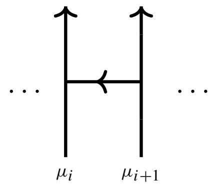
- The generator f_i acts by stacking the web

(31)

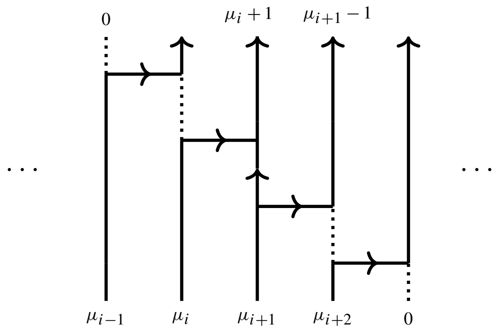


on the top of W_D. This means that f_i acts on $\mathcal{R}^\Lambda(n, d)$ as the functor that adds a strand labeled i to the left of D.

- To define the action of e_i we stack the web



on the top of W_D, then we use Tubbenhauer's trick (30) to put in a form that uses only F's. The transformation in (30) is not local and in order to be well defined one needs to keep trace of the indices before and after acting with an e_i. Tubbenhauer's trick gives

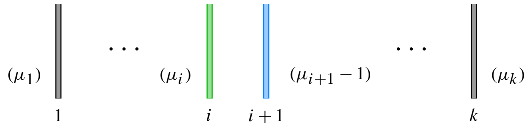


Every time we act with an e_i we embed $U_q(\mathfrak{gl}_k) \hookrightarrow U_q(\mathfrak{gl}_{k+1})$ and set

$$e_i(W_D) = f_{1^{(\mu_1)}\dots i-1^{(\mu_{i-1})}} f_i^{(\mu_i)} f_{i+1}^{(\mu_{i+1}-1)} f_{i+2^{(\mu_{i+2})}\dots k^{(\mu_k)}}(\mu, 0)(W_D).$$

After being acted with an e_j, f_i acts on W_D through the web corresponding to $f_{i+1}(\mu, 0)$.

We define the action of e_i on $\mathcal{R}^\Lambda(k, d)$ as the superfunctor that adds



to the left of D (here (μ_1), etc., are the thicknesses) that is, we act with the identity 2-morphism of $F_{1^{(\mu_1)}\dots i-1^{(\mu_{i-1})}} F_i^{(\mu_i)} F_{i+1}^{(\mu_{i+1}-1)} F_{i+2^{(\mu_{i+2})}\dots k^{(\mu_k)}}(\mu, 0)$.

Denote $\Phi(e_i)$ and $\Phi(f_i)$ the morphisms in $\mathfrak{R}^\Lambda$ that act as endofunctors of $\mathcal{R}^\Lambda(n, d)$ through the action above. It is clear that $\Phi(uv) = \Phi(u)\Phi(v)$ for u, $v \in U_q(\mathfrak{gl}_k)$. Note that $\Phi(1)(\mu)$ is a canonical element $F_{\text{can}}(\mu)$ as introduced in (23).

Lemma 4.1. *We have natural isomorphisms*

$$\Phi(e_i)\Phi(f_i)(\lambda) \simeq \Phi(f_i)\Phi(e_i)(\lambda) \oplus \Phi(1)^{\bigoplus[\bar{\lambda}_i]}(\lambda) \qquad \text{if } \bar{\lambda}_i \geq 0,$$
$$\Phi(f_i)\Phi(e_i)(\lambda) \simeq \Phi(e_i)\Phi(f_i)(\lambda) \oplus \Phi(1)^{\bigoplus[-\bar{\lambda}_i]}(\lambda) \quad \text{if } \bar{\lambda}_i \leq 0.$$

Proof. These are instances of the categorified higher Serre relations. Denote $F_u = F_{1^{(\lambda_1)}\dots i-1^{(\lambda_{i-1})}}$ and $F_d = F_{i+2^{(\lambda_{i+2})}\dots k^{(\lambda_k)}}$. We have

$$\Phi(e_i)\Phi(f_i)(\lambda) = F_u F_i^{(\lambda_i-1)} F_{i+1}^{(\lambda_{i+1})} F_d F_i(\lambda, 0)$$
$$\simeq F_u F_i^{(\lambda_i-1)} F_{i+1}^{(\lambda_{i+1})} F_i(\dots, \lambda_i, \lambda_{i+1}, 0, \lambda_{i+2}, \dots) F_d, (\lambda, 0),$$

and $\quad \Phi(f_i)\Phi(e_i)(\lambda) = F_t F_{i+1} F_i^{(\lambda_i)} F_{i+1}^{(\lambda_{i+1}-1)} F_b(\lambda, 0),$

and therefore, it is enough to check that the relations above are satisfied by the superfunctors $F_i^{(\lambda_i-1)} F_{i+1}^{(\lambda_{i+1})} F_i(\lambda_i, \lambda_{i+1}, 0)$ and $F_{i+1} F_i^{(\lambda_i)} F_{i+1}^{(\lambda_{i+1}-1)}(\lambda_i, \lambda_{i+1}, 0)$. Suppose $\lambda_i \geq \lambda_{i+1}$. Then we have $\lambda_i \in \{1, 2\}$ and $\lambda_{i+1} \in \{0, 1\}$. The computations involved are rather simple and we can check the four cases separately.

(1) $(\lambda_i, \lambda_{i+1}) = (1, 0)$:

$$\Phi(e_i)\Phi(f_i)(\lambda) = F_i^{(\lambda_i-1)} F_{i+1}^{(\lambda_{i+1})} F_i(\lambda_i, \lambda_{i+1}) = F_i(1, 0) = 0 \oplus F_{\text{can}}(1, 0),$$
$$= \Phi(f_i)\Phi(e_i)(\lambda) \oplus \Phi(1)(\lambda).$$

(2) $(\lambda_i, \lambda_{i+1}) = (1, 1)$:

$$\Phi(e_i)\Phi(f_i)(\lambda) = F_i F_{i+1}(1, 1, 0) = \Phi(f_i)\Phi(e_i)(\lambda).$$

(3) $(\lambda_i, \lambda_{i+1}) = (2, 0)$:

$$\Phi(e_i)\Phi(f_i)(\lambda) = F_i F_i(2, 0, 0)$$
$$\simeq q F_i^{(2)}(2, 0, 0) + q^{-1} F_i^{(2)}(2, 0, 0) = \Phi(1)^{\oplus[2]}(\lambda).$$

(4) $(\lambda_i, \lambda_{i+1}) = (2, 1)$:

$$\Phi(e_i)\Phi(f_i)(\lambda) = F_i F_{i+1} F_i(2, 1, 0)$$
$$\simeq 0 \oplus F_i^{(2)} F_{i+1}(2, 1, 0) = \Phi(f_i)\Phi(e_i)(\lambda) \oplus \Phi(1)(\lambda).$$

An this proves the first isomorphism in the statement. The second isomorphism can be checked using the same method. □

The proof of Lemma 4.1 uses several supernatural transformations between the various compositions of $\Phi(f_i)(\lambda)$ and $\Phi(e_i)(\lambda)$ and $\Phi(1)(\lambda)$ that can be given a presentation in terms of the diagrams from $\mathfrak{R}$. We act with such diagrams by stacking them on the top of the diagrams for the image of Φ. On the weight space $(1, 1)$ these maps coincide with the maps used to define the chain complex for a tangle diagram in the previous section. In the general case these maps are units and counits of adjunctions in the following.

Lemma 4.2. *Up to degree shifts, the functor $\Phi(e_i)$ is left and right adjoint to $\Phi(f_i)$.*

Lemma 4.3. *We have the following natural isomorphisms*:

$$\Phi(e_j)\Phi(f_i)(\lambda) \simeq \Phi(f_i)\Phi(e_j)(\lambda) \qquad \textit{for } i \neq j,$$
$$\Phi(f_i)\Phi(f_{i\pm1})\Phi(f_i)(\lambda) \simeq \Phi(f_i^{(2)})\Phi(f_{i\pm1})(\lambda) \oplus \Phi(f_{i\pm1})\Phi(f_i^{(2)})(\lambda),$$
$$\Phi(e_i)\Phi(e_{i\pm1})(\lambda)\Phi(e_i) \simeq \Phi(e_i^{(2)})\Phi(e_{i\pm1})(\lambda) \oplus \Phi(e_{i\pm1})\Phi(e_i^{(2)})(\lambda).$$

Proof. The proof consists of a case-by-case computation. We illustrate the proof with the case of $\Phi(e_i)\Phi(f_{i+1})(\lambda) \simeq \Phi(f_{i+1})\Phi(e_i)(\lambda)$ and leave the rest to the reader. We have

$$\Phi(e_i)\Phi(f_{i+1})(\lambda) = F_i^{(\lambda_i)} F_{i+1}^{(\lambda_{i+1}-2)} F_{i+2}^{(\lambda_{i+2}+1)} F_{i+1}(\lambda),$$
$$\text{and} \quad \Phi(f_{i+1})\Phi(e_i)(\lambda) = F_i^{(\lambda_i)} F_{i+2} F_{i+1}^{(\lambda_{i+1}-1)} F_{i+2}^{(\lambda_{i+2})}(\lambda),$$

which are zero unless $\lambda_{i+1} = 2$ and $\lambda_{i+2} \in \{0, 1\}$. If $\lambda_{i+1} = 2$ these can be written

$$\Phi(e_i)\Phi(f_{i+1})(\lambda) = F_i^{(\lambda_i)} F_{i+2}^{(\lambda_{i+2}+1)} F_{i+1}(\lambda),$$
$$\text{and} \quad \Phi(f_{i+1})\Phi(e_i)(\lambda) = F_i^{(\lambda_i)} F_{i+2} F_{i+1} F_{i+2}^{(\lambda_{i+2})}(\lambda).$$

The case $\lambda_{i+2} = 0$ is immmediate and the case $\lambda_{i+2} = 1$ follows from the Serre relation (8)–(9). □

As explained in [Brundan and Ellis 2017, Sections 1.5 and 6] the Grothendieck group of a ($\mathbb{Z}$-graded) monoidal supercategory is a $\mathbb{Z}[q^{\pm1}, \pi]/(\pi^2-1)$-algebra. Nontrivial parity shifts will occur when applying Tubbenhauer's trick. All the above can be used to prove the following.

Theorem 4.4. *The assignment above defines an action of $U_q(\mathfrak{gl}_k)$ on $\mathcal{R}^\Lambda(k, d)$. With this action we have an isomorphism of $K_0(\mathcal{R}^\Lambda(k, d))$ with the irreducible, finite-dimensional, $U_q(\mathfrak{gl}_k)$-representation of highest weight Λ at $\pi = 1$.*

Acknowledgements

We thank Daniel Tubbenhauer, Kris Putyra and Grégoire Naisse for interesting discussions. The author was supported by the Fonds de la Recherche Scientifique — FNRS under Grant no. MIS-F.4536.19.

References

[Brundan and Ellis 2017] J. Brundan and A. P. Ellis, "Monoidal supercategories", *Comm. Math. Phys.* **351**:3 (2017), 1045–1089. MR Zbl

[Ellis and Lauda 2016] A. P. Ellis and A. D. Lauda, "An odd categorification of $U_q(\mathfrak{sl}_2)$", *Quantum Topol.* **7**:2 (2016), 329–433. MR Zbl

[Ellis and Qi 2016] A. P. Ellis and Y. Qi, "The differential graded odd nilHecke algebra", *Comm. Math. Phys.* **344**:1 (2016), 275–331. MR Zbl

[Ellis et al. 2014] A. P. Ellis, M. Khovanov, and A. D. Lauda, "The odd nilHecke algebra and its diagrammatics", *Int. Math. Res. Not.* **2014**:4 (2014), 991–1062. MR Zbl

[Kang et al. 2013] S.-J. Kang, M. Kashiwara, and S.-j. Oh, "Supercategorification of quantum Kac–Moody algebras", *Adv. Math.* **242** (2013), 116–162. MR Zbl

[Kang et al. 2014] S.-J. Kang, M. Kashiwara, and S.-j. Oh, "Supercategorification of quantum Kac–Moody algebras, II", *Adv. Math.* **265** (2014), 169–240. MR Zbl

[Khovanov and Lauda 2009] M. Khovanov and A. D. Lauda, "A diagrammatic approach to categorification of quantum groups, I", *Represent. Theory* **13** (2009), 309–347. MR Zbl

[Khovanov and Lauda 2010] M. Khovanov and A. D. Lauda, "A categorification of quantum $\mathfrak{sl}(n)$", *Quantum Topol.* **1**:1 (2010), 1–92. MR Zbl

[Khovanov et al. 2012] M. Khovanov, A. D. Lauda, M. Mackaay, and M. Stošić, *Extended graphical calculus for categorified quantum* $\mathfrak{sl}(2)$, Mem. Amer. Math. Soc. **1029**, Amer. Math. Soc., Providence, RI, 2012. MR Zbl

[Lauda and Egilmez 2018] A. D. Lauda and I. Egilmez, "DG structures on odd categorified quantum $\mathfrak{sl}(2)$", preprint, 2018. arXiv

[Lauda and Russell 2014] A. D. Lauda and H. M. Russell, "Oddification of the cohomology of type A Springer varieties", *Int. Math. Res. Not.* **2014**:17 (2014), 4822–4854. MR Zbl

[Lauda et al. 2015] A. D. Lauda, H. Queffelec, and D. E. V. Rose, "Khovanov homology is a skew Howe 2-representation of categorified quantum $\mathfrak{sl}_m$", *Algebr. Geom. Topol.* **15**:5 (2015), 2517–2608. MR Zbl

[Lusztig 1993] G. Lusztig, *Introduction to quantum groups*, Progr. Math. **110**, Birkhäuser, Boston, 1993. MR Zbl

[Naisse and Vaz 2018] G. Naisse and P. Vaz, "Odd Khovanov's arc algebra", *Fund. Math.* **241**:2 (2018), 143–178. MR Zbl

[Ozsváth et al. 2013] P. S. Ozsváth, J. Rasmussen, and Z. Szabó, "Odd Khovanov homology", *Algebr. Geom. Topol.* **13**:3 (2013), 1465–1488. MR Zbl

[Putyra 2014a] K. K. Putyra, "A 2-category of chronological cobordisms and odd Khovanov homology", pp. 291–355 in *Knots in Poland, III: Part III* (Warsaw/Będlewo, 2010), edited by J. H. Przytycki and P. Traczyk, Banach Center Publ. **103**, Polish Acad. Sci. Inst. Math., Warsaw, 2014. MR Zbl

[Putyra 2014b] K. K. Putyra, *On a triply-graded generalization of Khovanov homology*, Ph.D. thesis, Columbia University, 2014, Available at https://search.proquest.com/docview/1537949284.

[Rouquier 2008] R. Rouquier, "2-Kac–Moody algebras", preprint, 2008. arXiv

[Shumakovitch 2011] A. N. Shumakovitch, "Patterns in odd Khovanov homology", *J. Knot Theory Ramifications* **20**:1 (2011), 203–222. MR Zbl

[Shumakovitch 2014] A. N. Shumakovitch, "Torsion of Khovanov homology", *Fund. Math.* **225**:1 (2014), 343–364. MR Zbl

[Stošić 2019] M. Stošić, "On extended graphical calculus for categorified quantum $\mathfrak{sl}(n)$", *J. Pure Appl. Algebra* **223**:2 (2019), 691–712. MR Zbl

[Tubbenhauer 2014] D. Tubbenhauer, "$\mathfrak{sl}_n$-webs, categorification and Khovanov–Rozansky homologies", preprint, 2014. arXiv

[Webster 2017] B. Webster, *Knot invariants and higher representation theory*, Mem. Amer. Math. Soc. **1191**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

Received December 16, 2019.

PEDRO VAZ
INSTITUT DE RECHERCHE EN MATHÉMATIQUE ET PHYSIQUE (IRMP)
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
BELGIUM
pedro.vaz@uclouvain.be

Guidelines for Authors

Authors may submit articles at msp.org/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 308 No. 1 September 2020

0030-8730(202009)308:1;1-F