Pacific Journal of Mathematics

Volume 232 No. 1 September 2007

PACIFIC JOURNAL OF MATHEMATICS

http://www.pjmath.org

Founded in 1951 by

E. F. Beckenbach (1906–1982)   F. Wolf (1904–1989)

See inside back cover or www.pjmath.org for submission instructions.

Regular subscription rate for 2007: $425.00 a year (10 issues). Special rate: $212.50 a year to individual members of supporting institutions. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. Prior back issues are obtainable from Periodicals Service Company, 11 Main Street, Germantown, NY 12526-5635. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

The Pacific Journal of Mathematics (ISSN 0030-8730) at the University of California, c/o Department of Mathematics, 969 Evans Hall, Berkeley, CA 94720-3840 is published monthly except July and August. 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.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS
at the University of California, Berkeley 94720-3840
A NON-PROFIT CORPORATION
Typeset in LaTeX

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

SYSTEMS OF BANDS IN HYPERBOLIC 3-MANIFOLDS

BRIAN H. BOWDITCH

Let M be a hyperbolic 3-manifold admitting a homotopy equivalence to a compact surface Σ, where the cusps of M correspond exactly to the boundary components of Σ. We construct a nested system of bands in M, where each band is homeomorphic to a subsurface of Σ times an interval. This band system is shown to have various geometrical properties, notably that the boundary of any Margulis tube is mostly contained in the union of the bands. As a consequence, one can deduce the result (conjectured by McMullen and proven by Brock, Canary and Minsky) that the thick part of the convex core of M has at most polynomial growth. Moreover the degree is at most minus the Euler characteristic of Σ. Other applications of this construction to the curve complex of Σ will be discussed elsewhere. The complex is related to the block decomposition of M described by Minsky, in his work towards Thurston's Ending Lamination Conjecture.

Introduction

This paper is primarily concerned with the geometry of hyperbolic 3-manifolds that are topologically products of a surface with the real line. More precisely, let M be a complete hyperbolic orientable 3-manifold admitting a homotopy equivalence $\chi : M \to \Sigma$ to a compact surface Σ. We assume that χ is "type preserving" in the sense that each boundary curve in Σ corresponds to a parabolic cusp in M. (We can allow for "accidental parabolics"; that is, parabolics in M need not be peripheral in Σ.) It follows from [Bonahon 1986] that M is homeomorphic to $\operatorname{int} \Sigma \times \mathbb{R}$. Manifolds of this sort have been intensively studied, for example in relation to Thurston's Ending Lamination Conjecture. (By lifting to an appropriate cover one can effectively reduce, at least in the indecomposable case, to manifolds of this type.)

The purpose of this paper is to describe a "band decomposition" of M, which captures much of its geometry. It gives a means of cutting the manifolds into simpler pieces, which can be understood intrinsically according to some inductive principle, and then fitted back together. One specific application is to give another

MSC2000: 57M50.
Keywords: 3-manifolds, hyperbolic geometry.

proof of the conjecture of McMullen that the thick part of such a manifold grows at most polynomially (see [Brock et al. 2004]), and give sharp bounds on the degree. Our polynomials are, in principle, algorithmically computable. Another application is to the geometry of the curve complex. One can show, for example, that the action of the mapping class group on the curve complex is acylindrical, and that stable lengths are uniformly rational. This is described in [Bowditch 2003]. Other applications of this work in turn show that the curve complex has finite asymptotic dimension [Bell and Fujiwara 2005] and has Yu's "property A" [Kida 2005]. It thus provides an example of hyperbolic 3-manifolds techniques being used to solve essentially combinatorial problems.

The decompositions described here have close links with Thurston's Ending Lamination Conjecture. As observed earlier, the indecomposable case can be essentially reduced to studying such manifolds: see [Minsky 2002; Brock et al. 2004; Bowditch 2005b]. (For adaptations of these ideas to the decomposable case see [Brock et al. $\geq$ 2007] and [Bowditch 2005a].) The key to this is to relate the geometry of M to the geometry of a "model" manifold constructed combinatorially. In principle a similar band decomposition could be constructed in the combinatorial model and then transferred to M. (Some discussion on how this may be achieved is given in [Bowditch 2005b].) However, such an approach is very indirect, and does not give a-priori computable constants. (At present, all known approaches to the Ending Lamination Conjecture involve limiting arguments, or equivalent, at some point.) Here we work directly from the 3-manifold, by a method that is, in principle, effective. This work is logically independent of the work on the Ending Lamination Conjecture cited above. We remark that another decomposition of M, which appears to be related, is discussed in [Soma 2003], and used there to study geometric limits of manifolds of this type.

1. Overview and examples

We start with an informal overview of what we mean by a "band system" and the properties we expect of it. These will be expressed more formally in Section 2.

We begin with the case of a compact surface, Σ, and a hyperbolic 3-manifold, M, without cusps, which admits a homotopy equivalence to Σ. To simplify the exposition we assume everything to be orientable. Thus, by [Bonahon 1986], M is homeomorphic to $\Sigma \times \mathbb{R}$. Its convex core, core(M), is homeomorphic to $\Sigma \times I$, where $I \subseteq \mathbb{R}$ is connected. In the geometrically finite case, core(M) and I are compact. We refer to the first and second coordinates as the horizontal and vertical directions respectively. There is no canonical homeomorphism, and so most statements in this section should be qualified with the phrase "after choosing suitable coordinates". In Section 2, we give a topological, coordinate-free means of expressing these ideas. In particular, we will define a "fibre" as an embedded closed surface whose inclusion in M is a homotopy equivalence. It is shown in [Brown 1966] that this has the form $\Sigma \times \{t\}$ in a suitable coordinate system.

A simple case is that of bounded geometry, i.e. where the injectivity radius is bounded below. If that happens, then the horizontal fibres $\Sigma \times \{t\}$ (in suitable coordinates), will have bounded diameter for all $t \in I$. In other words, we can foliate the convex hull with bounded diameter surfaces. In the general case, however, we get a set of disjoint short closed curves. These are unlinked, i.e. each has the form $\alpha \times \{t\}$ for a closed curve α in Σ [Otal 1995; 2003]. Any such curve will be the core of a Margulis tube. This time, the fibres can be taken to intersect the tubes in annuli, and such that the diameter of each component, after removing the tubes, is bounded. This controls the geometry in the horizontal direction. However there is no natural way of choosing vertical coordinates. For example, two fibres $\Sigma \times \{t\}$ and $\Sigma \times \{u\}$ may be close together on one side of a Margulis tube, but far apart on the other; and there might be no choice of coordinate system that will remedy this consistently. This is the kind of phenomenon the band system is designed to come to grips with.

We can also bring rank-one cusps into the picture. In this case, we allow Σ to be a closed surface with boundary. By hypothesis, each boundary component corresponds to a cusp of M. On removing these cusps, we get a manifold homeomorphic to $\Sigma \times \mathbb{R}$, and a similar discussion applies to this space. We may also get "accidental" cusps — homotopic to nonperipheral simple closed curves of Σ. These accidental cusps play a similar role to Margulis tubes. For the purposes of exposition, we will ignore accidental cusps in the discussion in this section.

It may happen that the boundary of each Margulis tube has bounded area. (This is necessarily the case if Σ is a one-holed torus or four-holed sphere; see [Minsky 1999].) In such a case, our band system will be empty. In general, however, one would expect these areas to be unbounded. Such tubes will form the anchors of a system of bands. A "band" is a subset of M of the form $\Phi \times J$ where Φ is a proper subsurface of Σ, and J is a compact subinterval of I. Each vertical boundary component, $\partial\Phi \times J$, is assumed to lie in the boundary of a Margulis tube. The band may intersect other tubes in solid tori. We should think of bands being long in the vertical direction, and narrow in the horizontal direction — that is narrow modulo the intersections with tubes (which are deemed not to contribute to the width). Qualitatively, a band, B, has similar geometry to that of the convex core of a geometrically finite manifold, N, with base surface Φ. Here, the tubes which meet the vertical boundary components of B should be thought of has having been "opened out" to rank-one cusps on N. This idea forms the basis of various inductive procedures, where we carry out induction on the complexity of the base surface. The induction starts with one-holed tori and four-holed spheres — there are no three-holed sphere bands.

Our goal will be to construct a system, $\mathcal{A}$, of disjoint bands with a number of geometric properties. Notably, we want the boundary of each Margulis tube to lie mostly inside the bands. More precisely, for each tube, the area lying outside the union of the bands is uniformly bounded in term of the complexity of Σ.

We can go on to construct a similar system inside each individual band, and then proceed inductively all the way down to one-holed tori and four-holed spheres, so as to give us a nested system, $\mathcal{B}$, of bands. In practice, it is this system we construct first. We can recover $\mathcal{A}$, if we want, as the set of outermost bands of $\mathcal{B}$.

The basic idea behind the construction of $\mathcal{B}$ is fairly simple. If there exists a sufficiently long band, $\Phi \times J$, with any given base surface, Φ, we include in $\mathcal{B}$ such a band which is almost as long as possible. By "long" we mean long in the vertical direction, in an appropriate sense, and the qualification "almost" means that we need the band to have collars attached at each end, in order to prevent neighbouring bands from bumping into each other. Some slight modification may necessary in some situations to ensure that the bands are nested, but that is mainly a technical issue. Most of the work of the proof will be in verifying that the boundaries of Margulis tubes are mostly taken up by the bands, so that, in some sense, the combinatorics of the band system does indeed capture most of the large scale geometry of M.

For most of the paper, we will simplify the exposition by assuming that Σ is closed, that M has no cusps, and that M is doubly degenerate, i.e. $\mathrm{core}(M) = M$, so that $I = \mathbb{R}$. The adaptation to the general case is discussed in Section 8.

We finish this section by giving a couple of simple examples. Suppose that there is just one Margulis tube, T, homotopic to a curve, γ, in Σ. Suppose γ separates Σ. Let Φ_1 and Φ_2 be the components of the complement of a small open annular neighbourhood of γ. There are four combinatorial possibilities for $\mathcal{A}$, namely: $\varnothing$, $\{\Phi_1 \times J_1\}$, $\{\Phi_2 \times J_2\}$ and $\{\Phi_1 \times J_1, \Phi_2 \times J_2\}$, where J_1, J_2 are intervals (Figure 1). Each of the bands meets ∂T in a single annulus. If γ is nonseparating, the possibilities are $\varnothing$ or $\{\Phi \times J\}$, where Φ is the complement of an open neighbourhood of γ in Σ. In the last case, the band meets ∂T in two annuli. This last possibility adds some complications to the formal description of bands, but has no particular geometric significance.

In the above, we will have $\mathcal{B} = \mathcal{A}$. More generally it is possible that the bands of $\mathcal{A}$ may themselves contain tubes and smaller bands of $\mathcal{B}$. Moreover, there may be many bands meeting any given tube. The general picture can get very complicated combinatorially (Figure 2). (This figure should elongated in the vertical direction to give a more accurate geometrical impression.)

It follows from the work on the Ending Lamination Conjecture that, in the generic case, the band system will be nonempty. However, explicit examples are

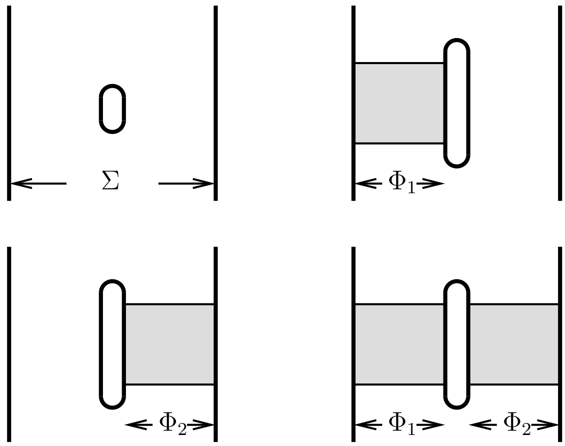


Figure 1

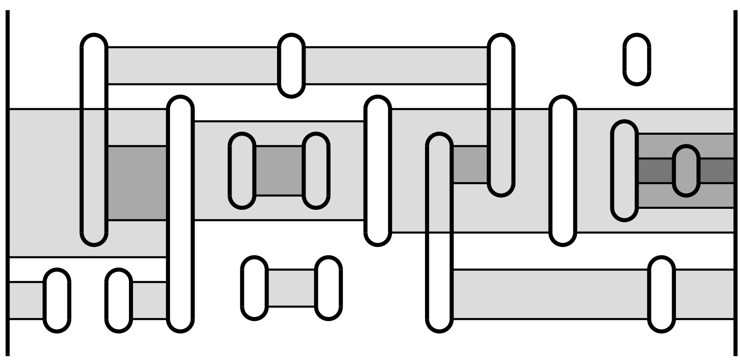


Figure 2

not so easy to construct and verify. Examples of product manifolds with no lower bound on injectivity radius were given in [Bonahon and Otal 1988]. Examples where the boundaries of Margulis tubes have arbitrarily large area (so that the band system is nonempty) were constructed in [Brock 2001].

2. Outline of results

In this section, we outline of the construction of the band decomposition, and summarise its main properties. We begin by recalling some standard facts.

For most of the paper, we will assume for simplicity that Σ is a closed orientable surface, and that M is orientable and has no cusps. Dealing with the general case will be mostly a matter of reinterpreting some of the definitions and constructions, as described in Section 8.

We know by tameness [Bonahon 1986] that M is homeomorphic to $\Sigma \times \mathbb{R}$. A *fibre* of M is an embedded surface homotopic (hence isotopic) to $\Sigma \times \{0\}$. A curve or subsurface is *unknotted* if it can be embedded in a fibre. More generally, a disjoint locally finite collection of embedded surfaces is *unlinked* if they can be simultaneously embedded in a collection of disjoint fibres.

Our discussion depends on certain "Margulis constants", η_0, η_1 etc. The Margulis Lemma tells us that there is some $\eta_0 > 0$ such that any closed geodesic, γ, of length at most η_0 in M is embedded, or finitely covers an embedded geodesic. Indeed, assuming it is primitive, it is the core of a "Margulis tube". Such a tube, T, is a solid torus, whose boundary, ∂T, is intrinsically euclidean. It comes equipped with a homotopically well defined meridian (bounding a disc in T). Otal [1995; 2003] shows that (provided η_0 is chosen small enough in relation to genus(Σ)), then γ is unknotted in M. Thus, ∂T also comes with a longitude (which can be homotoped to infinity in the complement of T). Such a longitude can also be described in terms of the framing of γ obtained by embedding it in a fibre. We can think of ∂T as foliated by euclidean geodesic longitudes of equal length. It turns out that this length is bounded above and below in terms of genus(Σ) (and the Margulis constant). This gives us a convenient normalisation: we fix a suitable $\eta > 0$ and write $T(\gamma) = T(\gamma, \eta)$ for the unique Margulis tube about γ whose longitudes all have length η. Provided η is small enough such tubes will be embedded and disjoint. We choose some other $\eta_1 > 0$ and let $\mathcal{T}$ be the set of all Margulis tubes, $T(\gamma, \eta)$ for which the core curve γ has length at most η_1. If $T \in \mathcal{T}$, we write $L(\partial T)$ for the "vertical length" of ∂T, i.e. the length of the circle obtained by collapsing each longitude to a point. (In other words, ∂T has area $\eta L(\partial T)$.) It turns out that $L(\partial T)$ is bounded away from 0, but there is no upper bound in general. The point of the band decomposition is that most of the vertical length of such a torus lies inside the union of the bands.

We write $\Theta(M)$ for the closure of $M \setminus \bigcup \mathcal{T}$ — the "thick part" of M. We equip $\Theta(M)$ with the induced path metric, d.

Definition. A *horizontal surface* in M is an unknotted surface, F, such that F meets each $T \in \mathcal{T}$, if at all, either in a single annulus whose boundary is precisely $F \cap \partial T$, or else in one or two (euclidean geodesic) longitudes of ∂T, both of which are boundary curves of F. Moreover, each boundary curve of F is a longitude of some element of $\mathcal{T}$.

We write $\mathcal{T}_I(F) \subseteq \mathcal{T}$ for the set of tubes meeting F in annuli.

Note that, under χ, F determines a homotopy class of subsurface of Σ, which we denote by $\phi(F)$.

Definition. We say that two horizontal surfaces F, G are *parallel* if they are disjoint and $\phi(F) = \phi(G)$.

Definition. A *band*, A, in M is a subset of M homeomorphic to $\Phi \times [-1, 1]$, where Φ is a proper subsurface of Σ, whose *horizontal* boundary, $\partial_H A = \Phi \times \{-1, 1\}$ consists of two horizontal surfaces (necessarily parallel) and whose *vertical boundary*, $\partial_V A = \partial\Phi \times [-1, 1]$ is a disjoint union of annuli, each lying in the boundary of some Margulis tube.

We denote the horizontal boundary components of A by $\partial_- A$ and $\partial_+ A$. (There is a canonical choice.) Any two parallel horizontal surfaces determine a band, A, with $\{\partial_- A, \partial_+ A\} = \{F, G\}$. We write $A = \langle F, G\rangle$. Write $\phi(A) = \phi(\partial_- A) = \phi(\partial_+ A)$.

Let $\mathcal{T}_I(A) \subseteq \mathcal{T}$ be the set of Margulis tubes completely contained in A. Each other tube meets A, if at all, in one or two vertical annuli, or else in a subsolid torus bounded by annuli of the form $T \cap \partial_- A$ or $T \cap \partial_+ A$ (either one of which may be empty).

Definition. The *width*, $W(F)$, of a horizontal surface, F, is the maximum diameter of a component of $F \cap \Theta(M)$ as measured in the path-metric, d, on $\Theta(M)$.

The *width*, $W(A)$, of a band, A, is defined as $W(A) = \max\{W(\partial_+ A), W(\partial_- A)\}$.

In some ways, it would be more natural to define "width" in terms of intrinsic diameter in the surface (in the induced path-metric) rather than using the ambient diameter in M. The problem is that our topological constructions will make it difficult to control intrinsic diameter, whereas the fact the that ambient diameters remain bounded is elementary.

Let $\mathcal{A}$ be a collection of bands in M. We write $\bigcup \mathcal{A}$ for their union. Given a Margulis tube, T, we write $L(\partial T, \mathcal{A})$ for the total vertical length of the union of annuli $\partial T \setminus \bigcup \mathcal{A}$.

In the discussion that follows, properties (A1), (A2), (A3), (A5), (A6) and (A9) will be proved later in the paper. Properties (A4), (A7) and (A8) are simple consequences, or can be assumed without loss of generality.

We shall show:

Theorem 0. *There are constants*, W_0, L_0, *depending on the topological type of* Σ (*and choice of Margulis constants*) *such that for any hyperbolic* 3*-manifold with a homotopy equivalence to* Σ, *we can find a collection*, $\mathcal{A}$, *of bands satisfying*:

(A1) *The elements of* $\mathcal{A}$ *are disjoint.*

(A2) *For each* $A \in \mathcal{A}$, $W(A) \le W_0$.

(A3) *For each* $T \in \mathcal{T}$, $L(\partial T, \mathcal{A}) \le L_0$.

We will see later that the bound on width in (A2) means that every point of A lies in a fibre of bounded width. As we will discuss below, we can strengthen (A1) to control the minimum distance between distinct bands, but at the cost of increasing the constant L_0 of (A3).

Note that if two bands, A and B are parallel (i.e. $\phi(A)=\phi(B)$) then they bound a third parallel band C. Thus A and B can be replaced by $A \cup C \cup B$. We see that there is no loss in assuming, in addition, that:

(A4) *No two distinct elements of $\mathcal{A}$ are parallel.*

We also note that the bands can all be assumed to lie in the convex core of M. (See the discussion of the "general case" below.)

There are a number of refinements we can make to Theorem 0.

Suppose A is a band. We write $\mathcal{T}_0(A) = \mathcal{T}_I(A) \cup \mathcal{T}_I(\partial_- A) \cup \mathcal{T}_I(\partial_+ A)$. The *exterior length*, $l(\pi, A)$ of a path π in A is the total (rectifiable) length of $\pi \setminus \bigcup \mathcal{T}_0(A)$.

Definition. The *height* $H(A)$ of a band A is the infimum (in fact minimum) exterior length of any path in A connecting $\partial_- A$ to $\partial_+ A$.

In other words, $H(A)$ is the shortest distance we need to travel to get across A, where travelling in the Margulis tubes (other than those that contain the vertical boundaries of A) costs us nothing.

We want a more quantitative way of saying that bands of $\mathcal{A}$ are disjoint, in fact a bounded distance apart. This can be expressed using the notion of a "collar".

Definition. If A is a band, a *top* (respectively *bottom*) *collar* of A is a band meeting A precisely in $\partial_- A$ (respectively $\partial_+ A$).

In other words, it has the form $\langle F, \partial_+ A \rangle$ or $\langle F, \partial_- A \rangle$, where F is a parallel horizontal surface.

Note that if A_+ and A_- are top and bottom collars of A, then $\hat{A} = A_+ \cup A \cup A_-$ is another band containing A. We refer to A, or more precisely, the pair $(A, \hat{A})$ as a *collared band*. Given $h \geq 0$, we say that A is *h-collared* if it admits a collar so that $H(A_+) \geq h$ and $H(A_-) \geq h$.

Addendum to Theorem 0. *There is some W_0 depending on the topological type of Σ such that given any H_0, $H_1 \geq 0$, we can find L_0 (depending on H_0, H_1 and the type of Σ), so that we can find a system of bands, $\mathcal{A}$, satisfying* (A1)–(A4) *above, together with*:

(A5) *Each band of $\mathcal{A}$ is H_0-collared.*

(A6) *If $A \in \mathcal{A}$, then $H(A) \geq H_1$.*

We can also assume if we want that $W(\hat{A}) \leq W_0$ for all $A \in \mathcal{A}$. By choosing $H_1 > 0$, we can assume that for each band, A, $A \cap \Theta(M)$ is connected (see the discussion of "primitive" bands in Section 3).

We shall see (Lemma 4.5) that $H(A)$ is uniformly bounded whenever $\phi(A)$ is a 3-holed sphere. Thus, by choosing H_0 or H_1 large enough, we can assume in addition:

(A7) *If $A \in \mathcal{A}$, then $\phi(A)$ is not a 3-holed sphere.*

Putting together (A3) and (A5), we see that there must be a bound on the number of components of $\partial T \setminus \bigcup \mathcal{A}$ for any $T \in \mathcal{T}$. This must in turn be at least the number of bands that meet ∂T. We deduce:

(A8) *There is some N_0 such that for all $T \in \mathcal{T}$, at most N_0 elements of $\mathcal{A}$ meet ∂T.*

Here N_0 depends on the topological type of Σ.

Finally, by choosing H_0 and/or H_1 large enough, we can ensure that our bands satisfy a topological property (defined in Section 2), namely:

(A9) *The elements of $\mathcal{A}$ are unlinked in M.*

As we have stated it, Theorem 0 says nothing about the intrinsic geometry of the bands. However, one could apply a similar construction to the interior of each band (compare the discussion of the general case below). Altogether, this would give us a larger system of bands, say $\mathcal{B}$, which are nested (see (A1′) below), rather than disjoint (as was required by (A1)), but which in addition satisfies a relative version of (A3), namely:

(A3′) *For each $B \in \mathcal{B}$ and $T \in \mathcal{T}_0(B)$ we have $L(\partial T \cap B, \mathcal{B}(B)) \leq L_0$.*

Here $\mathcal{B}(B) \subseteq \mathcal{B}$ is the set of bands strictly contained in B. In practice, we shall construct such a system $\mathcal{B}$ directly, and recover $\mathcal{A}$ as the set of outermost bands of $\mathcal{B}$.

There are some further refinements one can make to the band system $\mathcal{B}$.

Definition. Given $k > 0$, we say that two bands, A and B, are *k-nested* if one of the following three conditions holds: $N(A \cap \Theta(M), k) \subseteq B$, $N(B \cap \Theta(M), k) \subseteq A$ or $d(A \cap \Theta(M), B \cap \Theta(M)) \geq k$. They are *nested* if they are k-nested for some $k > 0$.

Here d is the path-metric on $\Theta(M)$, and $N(., k)$ denotes k-neighbourhood in $\Theta(M)$.

We can replace (A1) by:

(A1′) *The elements of $\mathcal{B}$ are nested.*

There is a final elaboration, alluded to earlier. Given any $H_2 > 0$, we can assume that the elements of $\mathcal{B}$ are H_2-nested. However, in this case, the constant L_0 of (A3′) will depend also on H_2.

The basic construction of the band system $\mathcal{B}$ is fairly simple. The constant W_0 is determined by the geometry of M (see Section 4.2). We choose some $H_4 \geq 0$ large enough in relation to H_0 and H_1. If A is a band with $W(A) \leq W_0$ and $H(A) \geq H_4$ then we choose such an A so as to maximise $H(A)$ among such bands with the same base surface $\phi(A)$. (Here we really mean "minimise" up to a small positive constant.) We can now find a subband, $B \subseteq A$, so that by setting $\hat{B} = A$, B is an H_0-collared band. By choosing H_0 large enough, we can ensure that any two such bands will be disjoint, at least modulo minor modification if one base surface should be contained in the other. We let $\mathcal{B}$ be the set of bands arising in this way. Most of the work is in verifying (A3). In fact, we will verify inductively a stronger version of (A3), starting with bands whose base surfaces have minimal complexity and working upwards to Σ. This procedure is discussed in Section 6.

In this section, we have only dealt explicitly with a special case. In general, we need to allow for parabolic cusps. One can also, in principle, account for the nonorientable case. Most of this will be outlined in Section 8. The main differences will be that in (A3) we should measure only vertical length in the convex core, but we can also allow for boundaries of accidental Margulis cusps. We may also need to allow for a finite number of "long bands" where one or more of the horizontal boundary components is at infinity.

We finally remark on the special case where Σ is a one-holed torus or four-holed sphere. This case is well analyzed in [Minsky 1999]. We know by (A7) that $\mathcal{B} = \varnothing$. Using (A3), we recover the fact that in such a manifold, the boundary of any Margulis tube has uniformly bounded vertical length, and hence bounded area.

3. The topology of M

First we consider band systems from a purely topological point of view. To simplify the exposition, we assume that Σ is a closed surface. (For the general case, see Section 8.)

Let X be the set of simple closed curves in Σ, defined up to homotopy. Unless otherwise stated, a subsurface, Φ, of Σ will be assumed to be connected, proper and essential (i.e. $\Phi \neq \varnothing$, $\Phi \neq \Sigma$ and Φ is not homotopic to a point). Indeed we shall normally assume that Φ is not an annulus, and that each boundary component of Φ is essential. (We allow for the complement of Φ in Σ to contain annular components.) We regard Φ as defined up to homotopy (or equivalently isotopy) in Σ. We write $\mathcal{F}$ for the set of (homotopy classes of) such surfaces. Given $\Phi \in \mathcal{F}$, we write $X(\Phi) \subseteq X$ for the set of curves that can be homotoped into Φ, and $X(\partial\Phi) \subseteq X(\Phi)$ for the set of homotopy classes of boundary curves. (Note that two curves in $\partial\Phi$ that bound an annular complementary component will get mapped to the same element of $X(\partial\Phi)$.)

Given $\Phi, \Psi \in \mathcal{F}$, we write $\Phi \subseteq \Psi$ to mean that Φ can be homotoped into Ψ. Note that this is equivalent to saying that $X(\Phi) \subseteq X(\Psi)$. A convenient way to imagine this would be fix any hyperbolic structure and identify the interior, $\operatorname{int}\Phi$, of Φ with an open subsurface with geodesic boundary. Such a realisation is unique. Moreover, $\Phi \subseteq \Psi$ in the sense above, if and only if their realisations satisfy $\operatorname{int}\Phi \subseteq \operatorname{int}\Psi$.

Definition. Given $\Phi, \Psi \in \mathcal{F}$, we say that $\Omega \in \mathcal{F}$ is a *component of* $\Phi \cap \Psi$ if we can homotope Φ, Ψ and Ω so that Ω is a connected component of $\Phi \cap \Psi$ in the usual sense.

The following is easily verified:

Lemma 3.1. *Suppose $\Phi, \Psi, \Omega \in \mathcal{F}$. Then Ω is a component of $\Phi \cap \Psi$ if and only if $X(\Omega) \subseteq X(\Phi) \cap X(\Psi)$ and $X(\partial\Omega) \subseteq X(\partial\Phi) \cup X(\partial\Psi)$.* □

Given $\Phi \in \mathcal{F}$, write $|\partial\Phi|$ for the number of boundary components. (This will be bigger than $|X(\partial\Phi)|$ whenever there is a complementary annular component.)

Definition. The *complexity*, $\kappa(\Phi)$, of Φ is defined by $\kappa(\Phi) = 3\,\mathrm{genus}(\Sigma) + |\partial\Phi| - 3$.

Note that if $\Phi \subseteq \Psi$, then $\kappa(\Phi) \le \kappa(\Psi)$, with equality only if $\Phi = \Psi$. Moreover, $\kappa(\Phi) = 0$ if and only if Φ is a 3-holed sphere.

Now let $M = \Sigma \times \mathbb{R}$, and let $\chi : M \to \Sigma$ be the projection map. We want to express various topological notions without making explicit reference to any coordinate system on M.

Definition. A *fibre* of M is the image of an injective homotopy equivalence of Σ to M.

It turns out (see [Brown 1966]) that any fibre is ambient isotopic to $\Sigma \times \{0\}$. Continuing inductively, we see that if $S_1, \dots, S_n$ are disjoint fibres, then $S_1 \cup \cdots \cup S_n$ has the form $\Sigma \times \{1, \dots, n\}$ up to isotopy (and permutation).

Definition. By an *unknotted* surface in M we mean a subsurface F of a fibre S, whose projection to Σ lies in $\mathcal{F} \sqcup \{\Sigma\}$.

This projection is well defined up to homotopy. We denote it by $\phi(F) \in \mathcal{F} \sqcup \{\Sigma\}$.

Definition. A collection of disjoint (unknotted) surfaces, $F_1, \dots, F_n$ is *unlinked* if there are disjoint fibres, $S_1, \dots, S_n$ with $F_i \subseteq S_i$ for each i.

One can generalise this to an infinite locally finite collection. In this case, the ambient fibres are disjoint, locally finite, and indexed by $\mathbb{N}$ or $\mathbb{Z}$.

We can extend these definitions to include closed curves in M (necessarily simple and essential in Σ). A collection of disjoint solid tori in M are said to be unlinked if their cores are unlinked. We define $\phi(\gamma) \in X$ and $\phi(T) \in X$ in the obvious way for an unknotted curve, γ, or solid torus, T.

As discussed in Section 2, if $T \subseteq M$ is an unknotted torus, then ∂T has a well defined meridian and longitude up to homotopy. (Together these generate $H_1(\partial T)$.)

Pushing surfaces. We describe a procedure for "pushing" one fibre off another to make them disjoint. We normally want to do this while fixing some subsurface or curve in the fibre. The main applications of this process (and its variants discussed later) will come in Section 7.

Suppose that S, $Z \subseteq M$ are fibres and that $F \subseteq Z$ is an essential surface or curve not meeting Z. We will produce a fibre, S', containing F, disjoint from Z, and contained in an arbitrarily small neighbourhood of $S \cup Z$.

We can assume that S and Z meet transversely. Let G_S be the closure of the component of $S \setminus Z$ containing F.

Step 1: We first arrange that each boundary curve of G_S is essential. For if not, start with a homotopically trivial boundary curve, $\alpha \subseteq \partial G_S \cap Z$, which is innermost in Z among such boundary curves. It bounds a disc, D_S, in S and a disc, D_Z, in Z. Since F is essential, we have $F \cap D_S = \varnothing$. Now replace D_S in S by D_Z pushed slightly off Z, and adjoin D_Z to G_S to get rid of the boundary curve α. We continue to perform such disc replacements until we rid ourselves of all such trivial boundary curves. Our new surface, S, may not be embedded. (It may intersect itself along certain trivial curves.) However it remains a homotopy equivalence, and G_S is embedded in M and still contains F. Moreover, $G_S \cap Z = \partial G_S$.

Step 2: Since each boundary curve of G_S is essential, there is a subsurface $G_Z \subseteq Z$ with $\partial G_Z = \partial G_S \subseteq M$ and with $\phi(G_Z) = \phi(G_S)$ (allowing for the possibility that G_S and G_Z are both annuli). There is thus a natural bijection between the components of $S \setminus G_S$ (as an immersed surface) and those of $Z \setminus G_Z$. We can thus replace each component of $S \setminus G_S$ with the corresponding component of $Z \setminus G_Z$, pushed slightly off Z. (Note that G_S is connected, and hence lies to one side of Z.) Since $Z \cap G_S = \partial G_S$, the resulting surface is embedded. It is clearly a homotopy equivalence, and hence a fibre containing F, as required.

Here is a simple consequence of the pushing process:

Lemma 3.2. *Suppose the $S_1, \dots, S_n$ are a set of fibres of M and for each i, $F_i \subseteq S_i$ is an unknotted surface or curve. If $F_i \cap S_j = \varnothing$ for all distinct i and j, then the surfaces, F_i, are unlinked in M.*

Proof. Assume inductively that we have disjoint fibres, $S'_1, \dots, S'_m$ with $F_j \subseteq S'_j$ for all $j \le m$, and $F_i \cap S'_j = \varnothing$ for all $i > m$. Now inductively push S_{m+1} off each of the fibres S'_j to obtain a fibre S'_{m+1} containing F_{m+1}, disjoint from each of the other S'_j, and contained in a small neighbourhood of $S_{m+1} \cup \bigcup_{j \le m} S'_j$. We see that $F_k \cap S'_{m+1} = \varnothing$ for all $k \ge m+2$. We eventually get the F_i lying in disjoint fibres as required. $\square$

Thick surfaces. A "thick surface" will give us a topological formulation of band.

Definition. An (unknotted) *thick surface*, A, in M is the image of an embedding of $\Phi \times [-1, 1]$ for some $\Phi \in \mathscr{F}$, such that $\Phi \times \{t\}$ is unknotted for some, hence every, $t \in [-1, 1]$.

We can assume that these surfaces map back to Φ under the projection χ. We write $\phi(A) = \Phi$. We refer to $\phi(A)$ as the *base surface* of A.

We can write $\partial A = \partial_H A \cup \partial_V A$, where $\partial_H A = \Phi \times \{-1, 1\}$ and $\partial_V A = \partial\Phi \times [-1, 1]$ are respectively the *horizontal* and *vertical* boundaries of A. Indeed we can write $\partial_H A = \partial_+ A \sqcup \partial_- A$, where $\partial_\pm A$ lies in a fibre $S_\pm$, where S_+ separates S_- from the positive end of M. One can check this is well-defined. By a *fibre* of A we mean the image of an injective homotopy equivalence of Φ into $A \setminus \partial_H A$, with $\partial\Phi = \Phi \cap \partial_V A$. As with M, a fibre of A, is isotopic in A to $\Phi \times \{0\}$.

Lemma 3.3. *Suppose $A \subseteq M$ is a thick surface and $F \subseteq M$ is an unknotted surface with $F \cap \partial_H A = \varnothing$. Let G be a nonannular component of $F \cap A$ meeting $\partial_V A$ only in essential (core) curves. Then $\phi(G)$ is a component of $\phi(F) \cap \phi(A)$.*

Proof. It is easy to see that $X(\phi(G)) \subseteq X(\phi(F)) \cap X(\phi(G))$ and $X(\partial\phi(G)) \subseteq X(\partial\phi(F)) \cup X(\partial\phi(A))$, and so the result follows by Lemma 3.1. $\square$

Corollary 3.4. *Suppose $A \subseteq M$ is a thick surface and $S \subseteq M$ is a fibre with $S \cap \partial_H A = \varnothing$. Suppose S meets each component of $\partial_V A$ if at all in a single core curve. Then $S \cap A$ is either empty or a fibre of A.*

Proof. If $S \cap A \neq \varnothing$, let G be a component of $S \cap A$. This cannot be an annulus. We apply Lemma 3.3 with $F = \Sigma$ to see that $\phi(G)$ is a component of $\phi(A)$, and hence equal to it. Thus the inclusion of G in A is a homotopy equivalence. Since $\partial G \subseteq \partial_V A$, it follows that G is a fibre. In particular, G meets each component of $\partial_V A$, and so $G = S \cap A$. $\square$

Definition. We say that a set of disjoint thick surfaces in *unlinked* if some (hence any) set of disjoint fibres thereof is unlinked.

Horizontal surfaces and bands. We now bring our topological Margulis tubes into play. Suppose that $\mathscr{T}$ is a locally finite disjoint collection of unlinked solid tori in M. There is a map $\phi : \mathscr{T} \to X$, which we assume to be injective. We also assume that for each $T \in \mathscr{T}$, ∂T comes equipped with a foliation by longitudes (referred to as *horizontal longitudes* if we need to clarify). We write $\Theta(M)$ for the closure of $M \setminus \bigcup \mathscr{T}$. For surfaces, the use the term "horizontal" to mean that it intersects the Margulis tubes nicely. More precisely:

Definition. A *horizontal surface* is an unknotted surface, $F \subseteq M$, such that there are two disjoint subsets $\mathscr{T}_\partial(F)$ and $\mathscr{T}_I(F)$ of $\mathscr{T}$ such that:

(1) For all $T \in \mathscr{T} \setminus (\mathscr{T}_I(F) \cup \mathscr{T}_\partial(F))$, $T \cap F = \varnothing$.

(2) For all $T \in \mathscr{T}_I(F)$, $T \cap F$ is an annulus whose boundary is precisely $\partial T \cap F$.

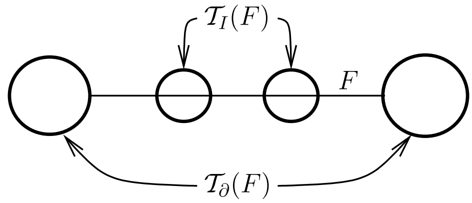


Figure 3

(3) For all $T \in \mathcal{T}_\partial(F)$, $T \cap F = T \cap \partial F$ consists of one or two horizontal longitudes.

(4) $\partial F \subseteq \bigcup \mathcal{T}_\partial(F)$.

(See Figure 3.)

Definition. A *horizontal fibre* is a horizontal surface that is also a fibre.

Clearly, if F is a horizontal fibre, then $\mathcal{T}_\partial(F) = \varnothing$. Otherwise, $F \in \mathcal{F}$.

Definition. Two horizontal surfaces are *parallel* if $\phi(F) = \phi(G)$ and $F \cap G = \varnothing$.

This implies that $\mathcal{T}_\partial(F) = \mathcal{T}_\partial(G)$.

Definition. A horizontal surface, F, is *primitive* if $\mathcal{T}_I(F) = \varnothing$.

Definition. A *piece* of a horizontal surface, F, is a connected component of $F \cap \Theta(M)$.

Note that a piece of F is a primitive horizontal surface. (Note also that $F \cap \Theta(M)$ might be connected even if F is not primitive.)

Next we come to the notion of a band. As discussed earlier, this a thick surface whose vertical boundary lies in the boundary of tubes. All other tubes, meet it, if at all, in solid tori. We need to allow for the possibility of a tube cutting all the way through a band, from the top to the bottom surface. If this doesn't happen, the band will be called "primitive". Here is a formal account.

Definition. A *band* is an unknotted thick surface, $B \subseteq M$, such that there are subsets $\mathcal{T}_\partial(B)$, $\mathcal{T}_I(B)$, $\mathcal{T}_+(B)$ and $\mathcal{T}_-(B)$ of $\mathcal{T}$ satisfying:

(1) The three sets $\mathcal{T}_\partial(B)$, $\mathcal{T}_I(B)$ and $\mathcal{T}_+(B) \cup \mathcal{T}_-(B)$ are mutually disjoint.

(2) If $T \in \mathcal{T} \setminus (\mathcal{T}_\partial(B) \cup \mathcal{T}_I(B) \cup \mathcal{T}_+(B) \cup \mathcal{T}_-(B))$ then $T \cap B = \varnothing$.

(3) If $T \in \mathcal{T}_I(B)$, then $T \subseteq B$ and $T \cap \partial_\pm B = \varnothing$.

(4) If $T \in \mathcal{T}_\partial(B)$, then $T \cap B = \partial T \cap B$ has one or two components, (each of) which is a component of $\partial_V B$ and lies between two horizontal longitudes of ∂T.

(5) $\partial_V B \subseteq \bigcup \mathcal{T}_\partial(B)$.

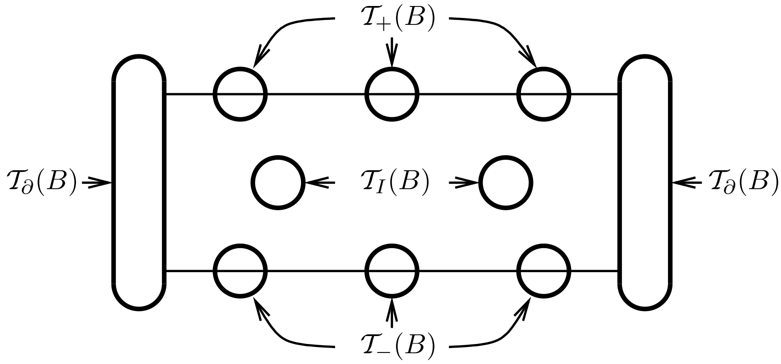


Figure 4

(6) If $T \in \mathcal{T}_+(B)$, then $T \cap \partial_+ B$ is an annulus whose boundary is $\partial T \cap \partial_+ B$ and consists of two horizontal longitudes of ∂T.

(7) As with (6) with $-$ instead of $+$.

(See Figure 4.)

Note that $\partial_\pm B$ is a horizontal surface, with $\mathcal{T}_\partial(\partial_\pm B) = \mathcal{T}_\partial(B)$ and $\mathcal{T}_I(\partial_\pm B) = \mathcal{T}_\pm(B)$. If $T \in \mathcal{T}_+(B) \cup \mathcal{T}_-(B)$, then T meets B in a subsolid torus. (Note that its complement in T will have two components if $T \in \mathcal{T}_+(B) \cap \mathcal{T}_-(B)$.)

We write $\mathcal{T}_0(B) = \mathcal{T}_I(B) \cup \mathcal{T}_+(B) \cup \mathcal{T}_-(B)$.

Clearly $\partial_+ B$ and $\partial_- B$ are parallel. Conversely, if F and G are parallel horizontal surfaces, then F and G determine a unique band, B, with $\{F, G\} = \{\partial_+ B, \partial_- B\}$. We write $B = \langle F, G \rangle$.

Definition. A band, B, is *primitive* if $\mathcal{T}_+(B) \cap \mathcal{T}_-(B) = \varnothing$.

Definition. A *piece* of a band B is the closure of a connected component of $B \setminus \bigcup(\mathcal{T}_+(B) \cap \mathcal{T}_-(B))$ (see Figure 5).

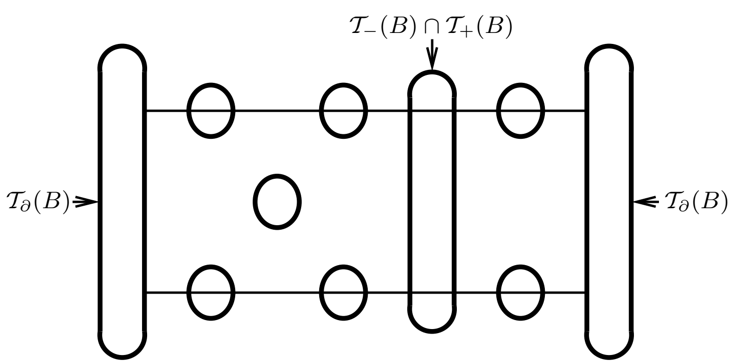


Figure 5

Note that a piece of a band is a primitive band.

Definition. A *horizontal fibre* of a band B is a fibre of B that is a horizontal surface.

We note that a fibre, F, of a band, B, divides F into two bands, namely $\langle F, \partial_- B\rangle$ and $\langle F, \partial_+ B\rangle$. We also note the following consequence of Lemma 3.3 and Corollary 3.4:

Lemma 3.5. *Let B be a band.*

(1) *If $F \subseteq M$ is a horizontal surface with $F \cap \partial_\pm B = \varnothing$ and G is a component of $F \cap B$, then G is a horizontal surface with $\phi(G) = \phi(F) \cap \phi(B)$.*
(2) *If S is a horizontal fibre with $S \cap B \neq \varnothing$ and $S \cap \partial_\pm B = \varnothing$, then $S \cap B$ is a horizontal fibre of B.* □

Pushing horizontal fibres. We need to elaborate on the pushing procedure described earlier, in order to take account of the positions of the tubes.

Suppose that S, Z are horizontal fibres, and that F is an essential surface or curve lying in some piece of F, with $F \cap Z = \varnothing$. As before, we want to "push" S off Z to obtain a fibre S' containing F. We need to refine our previous "pushing" procedure slightly in order to ensure that the resulting fibre is horizontal.

We can assume that S meets Z transversely. We can also assume that if $T \in \mathcal{T}_I(S) \cap \mathcal{T}_I(Z)$, then $\partial T \cap S \cap Z = \varnothing$, and that the annuli $S \cap T$ and $Z \cap T$, meet, if at all, in single core curve. Let G_S be the closure of the component of $S \setminus Z$ containing F. Thus, each boundary curve of G_S is either a core curve of some solid torus, or else lies in a piece of Z.

Step 1: First get rid of the homotopically trivial components of ∂G_S as before, noting that each of the discs, D_Z, lies in some piece of Z.

Step 2: Let $G_Z \subseteq Z$ be the subsurface with $\partial G_Z = \partial G_S$ and $\phi(G_S) = \phi(G_Z)$. Let S_1 be the surface obtained by replacing the components of $S \setminus G_S$ with the corresponding components of $Z \setminus G_Z$. As before, S_1 is a fibre containing F.

Step 3: We may need to adjust S so that it becomes horizontal. Suppose that $T \in \mathcal{T}$. Now $S_1 \cap T$ is empty or consists of one or two annuli (each of the form $S \cap T$ or $Z \cap T$ pushed slightly, or obtained by surgery on $S \cap T$ and $Z \cap T$ in the case where they intersect in a core curve.) Thus, the only thing that can go wrong is that we may have a torus, T, with $S_1 \cap T = P \sqcup Q$, where P, Q are annuli. These are homotopic in S_1 and hence bound a third annulus, $R \subseteq S_1$. Now if $S_1 \cap F = \varnothing$, then we can just push R into T so that $S_1 \cap T$ becomes a single annulus. After doing this a finite number of times, we obtain our horizontal fibre.

It remains to worry about the case where F meets, and hence is contained in R. Now we cannot have $G_S \subseteq R$ (otherwise the process of obtaining S_1 would not have produced any such double annuli). Nor can we have $R \subseteq G_S$ (since $G_S \subseteq S$, and we are assuming S to be horizontal). It follows that the annulus $V = G_S \cap R \subseteq F$ has one boundary component, α, in $T \cap S$, and the other boundary component, β, in $\partial G_S \subseteq Z$.

At this point, we forget about Step 2, and instead do:

Step 2′: Recall that we have $F \subseteq V \subseteq S$ with $V \cap T = \alpha$, $V \cap Z = \beta$, $\partial V = \alpha \sqcup \beta$ and $T \cap Z \neq \varnothing$. Now $T \cap Z$ and β bound an annulus, $W \subseteq Z$. Let γ be the boundary curve of $T \cap Z$ on the other side of $W \cap T$. We connect γ to α by an annulus $Y \subseteq T$, and now replace $(T \cap Z) \cup W$ in Z by $Y \cup V$. Pushing this surface slightly off Z, we get our desired horizontal fibre, S'.

We finally note that this pushing process can be applied to subsurfaces in the following sense.

Suppose that S is a fibre, and $F \subseteq S \cap \Theta(M)$ is an essential surface or curve. Suppose that $J \subseteq S$ is a horizontal surface containing F, and that K is another horizontal surface with $\phi(K) = \phi(J)$, and with $K \cap S \subseteq J$. We can form a fibre Z with $K \subseteq Z$ and with Z agreeing with S on all complementary horizontal surfaces. Now applying the procedure above, we see that S remains unchanged on the complement of K (modulo modifications in the solid tori containing boundary components of K). We have thus effectively pushed S off K, while retaining F unchanged.

4. Metric properties

So far, we have only considered the topological structure of M. In this section we summarise its key metric properties. We shall assume that M is (constant curvature) hyperbolic, though the essential points can be interpreted for more general metrics, for example, in pinched negative curvature.

Again, we assume that M has no parabolic cusps, and admits a homotopy equivalence $\chi : M \to \Sigma$ to a closed surface Σ. By [Bonahon 1986], M is homeomorphic to $\Sigma \times \mathbb{R}$. By [Otal 1995; 2003] the set $\mathcal{T}$ of Margulis tubes is unlinked. We write $Y = \mathrm{core}(M)$.

Recall that $\Theta(M)$ is the thick part of M, with induced path metric d. At least once the essential properties are derived, only the geometry on $\Theta(M)$ will be relevant to future discussion.

We note the following four geometric features of M.

4.1. *Geometry of tori.* We shall assume that the "thick part" of M is defined in such a way as to simplify the handling of constants. The standard definition of thick part involves fixing a sufficiently small Margulis constant, $\epsilon > 0$, and defining it to be the set of points where the injectivity radius is at least ϵ. In this way, the thin part is a disjoint union of tubes. However, we get a similar qualitative picture if we allow for different tubes to be defined by different injectivity radii, provided they range between two fixed positive constants. This allows us to make certain additional metric assumptions about the tubes that will simplify further discussion.

Suppose $T \in \mathcal{T}$. The geodesic core of a Margulis tube lies in the convex core, Y, of M. The boundary, ∂T, is a euclidean torus, foliated by geodesic longitudes. It meets ∂Y, if at all, in a collection of geodesic longitudes. In fact, T either lies in Y or meets Y is a solid torus bounded by one or two annuli.

It is convenient to assume that all geodesic longitudes of all ∂T they all have the same length, say η. This can be achieved by noting that every longitude in $\partial T \cap Y$ lies inside some horizontal fibre. (This follows from work of Otal; see the discussion in Section 4.2 below.) In general, its length will necessarily be bounded between two positive constants, and so, using the observation of the preceding paragraph, it can be assumed to be fixed. The constant, η, can be chosen to depend only on the complexity, $\kappa(\Sigma)$, of Σ (though could also be taken to be arbitrarily small). These geodesic longitudes are deemed to be horizontal. We can also assume that there is a lower bound on the distance between two such Margulis tubes, which we can also take to be η. We will also want to assume that the boundaries of a Margulis tube T has extrinsic curvature close enough to 1 (the extrinsic curvature of a horosphere). This can be achieved by assuming the length of the core geodesic is small in relation to η, so that it lies deep inside T. (Again, using the principle of the first paragraph.) Note that, by definition, there is some lower bound on the lengths of closed geodesics in the thick part, $\Theta(M)$. This depends on the Margulis constant, η, we have fixed, and the maximal lengths of core curves of tubes that we are allowing.

4.2. ***Horizontal fibres.*** There is some constant, W_0, depending on $\kappa(\Sigma)$ (and η) such that any point of $Y \cap \Theta(M)$ is contained in a horizontal fibre $S \subseteq Y$ of width $W(S) < W_0$. (Recall that $W(S)$ is defined as the maximal diameter of any piece of S measured in the path metric d on $\Theta(M)$.) In particular, any horizontal longitude of any torus is contained in such a surface. Note that by taking strict inequality, we can push such a surface slightly off itself to give a disjoint surface while maintaining the same bound.

This can be achieved using various standard arguments. The main ideas of the construction can be found in [Otal 1995; 2003]. We first need to use the fact that every point of M lies in the image of a uniformly lipschitz homotopy equivalence, $\phi : \Sigma \to M$, where Σ carries some hyperbolic metric. The usual argument for this is based on some form of interpolation of pleated surfaces; see [Thurston 1979]. A technically simpler approach is to use singular hyperbolic surfaces of the type described in [Bonahon 1986]. In particular, the "filling theorem" of [Canary

1996], gives us what we need. (The latter gives us a singular hyperbolic metric on Σ, but that works just as well.) Now the intrinsic diameter of each component of $\Sigma \cap \phi^{-1}\Theta(M)$ is bounded in terms of $\kappa(\Sigma)$ (and η). We can homotope ϕ a bounded distance so that the preimage of the set of tubes is a set of nonparallel essential annuli in Σ, whose boundary curves are horizontal. We can now perform a variant of the construction of [Freedman et al. 1983], as described in [Otal 2003], to give us an embedded surface, F, in an arbitrarily small neighbourhood of $\phi(\Sigma)$. (Some care is needed to ensure that the original longitude remains in F.) Now, the ambient diameter of each component of $F \cap \Theta(M)$ remains bounded. (In principle, one can achieve a bound on the intrinsic diameters of such components, but this would require more work.)

4.3. ***Bounded geometry.*** Since the injectivity radius of $\Theta(M)$ is bounded below, it has "bounded geometry". One way of exploiting this, following [Gromov 2007], is this. Let $r > 0$ be the lower bound on injectivity radius, as in Section 4.1, and assume that any pair of distinct tubes are distance at least $2r$ apart. A subset $V \subseteq Y \cap \Theta(M)$ is said to be r-separated if $d(x, y) \geq r$ for all distinct $x, y \in V$. We can form a graph, $\Delta(V)$, with vertex set V, and with $x, y \in V$ adjacent in $\Delta(V)$ if $d(x, y) \leq 3r$. Bounded geometry implies that the degree of any vertex of such a graph is uniformly bounded. We note that we could choose V so that $\Theta(M)$ lies in a $(2r)$-neighbourhood of V. Such a set is called an *r-net*. In this case, the "nerve", $\Delta(V)$, approximates distance in $\Theta(M)$ to within linear bounds.

From our choice of r, the r-ball about any point $x \in \Theta(M)$ a distance at least r from any tube will be isometric to an r-ball in hyperbolic 3-space. If x is close to a tube T, then it will have a piece of this tube removed, and slightly distorted geometry. (Since we are defining balls in terms of the metric d.) In any case, it is a nice contractible set.

4.4. ***Three-holed spheres.*** The following (while not really essential to the construction) will tell us that no band in our system has base surface a 3-holed sphere. (In retrospect, this explains why boundaries of Margulis tubes have bounded area in the case of a 1-holed torus or 4-holed sphere.)

Lemma 4.5. *There is a constant, $H_3 > 0$ such that if $B \subseteq M$ is a band with base surface, $\phi(B)$, a 3-holed sphere and with $W(B) \leq W_0$, then we can connect $\partial_+ B$ to $\partial_- B$ by a path in B of length at most H_3.*

Proof. Let $\mathcal{T}_\partial(B) = \mathcal{T}_0(B) = \{T_1, T_2, T_3\}$, and let $\gamma_i^\pm = T_i \cap \partial_\pm B$. There is a path, $\sigma_i^\pm$ in $\Theta(M)$ connecting $\gamma_i^\pm$ to $\gamma_{i+1}^\pm$ of length at most W_0 (taking indices mod 3). Since we are dealing with a 3-holed sphere, we see that each σ_i^+ is homotopic to σ_i^- rel $\partial T_i \cup \partial T_{i+1}$. Lifting this picture to $\mathbb{H}^3$, we get six paths, $\tilde{\sigma}_i^\pm$ connecting the three sets $\tilde{T}_i$, each of these sets being a uniform neighbourhood of a bi-infinite geodesic.

Simple hyperbolic geometry now gives us a bound on the distance between $\tilde{\sigma}_i^+$ and $\tilde{\sigma}_i^-$ in the boundary of $\tilde{T}_i$. Projecting back to M gives the result. □

We shall assume henceforth that we have fixed the constants η and r (depending on $\kappa(\Sigma)$). The constants W_0 and H_3 are thus determined.

We remark that there are other important properties of the geometry of M, for example the "Uniform Injectivity Theorem" for pleated surfaces (which seems central to the Ending Lamination Conjecture). However, we make no use of this here — which means that all our constants are, in principle, computable functions of $\kappa(\Sigma)$.

5. The band system

We now describe more carefully the construction of a nested band system, $\mathcal{B}$.

Definition. The *exterior length*, $l(\pi)$, of a path π in M is the rectifiable length of $\pi \cap \Theta(M)$.

Definition. A *vertical fibre* of a band, A, is a path in $A \setminus \partial_V A$ connecting $\partial_+ A$ to $\partial_- A$.

Definition. The height, $H(A)$, of a band, A, is the infimum of the exterior lengths of vertical fibres.

Note that A is primitive if and only if $H(A) > 0$. In fact, when this is positive it is more convenient to take $H(A)$ to be this infimum plus an arbitrarily small positive constant. Thus we can assume we have a vertical fibre of length at most $H(A)$.

Definition. Given $x \in A$, the *depth* of x in A, denoted $D(x, A)$ is the infimum of $l(\pi)$ as π varies over all paths in A connecting x to $\partial_H A$ in $A \setminus \partial_V A$.

If $Q \subseteq A$, we write $D(Q, A) = \inf\{D(x, A) \mid x \in Q\}$ for the *depth* of Q in A.

Again it is convenient to add a small positive constant, or to pretend that the infimum is attained.

Let $\nu = \nu(\Sigma)$ be minus the Euler characteristic. This is the number of 3-holed spheres in any pants decomposition of Σ. It thus bounds the number of pieces in any horizontal surface in M.

Lemma 5.1. *Suppose A is a band, $F \subseteq M$ is a horizontal surface, and $x \in F \cap A$ with $D(x, A) > \nu W(F)$. Let G be the component of $F \cap A$ containing x. Then G is a horizontal surface with $\phi(G)$ a component of $\phi(F) \cap \phi(A)$. In particular, if F is a horizontal fibre of M, then $F \cap A$ is a horizontal fibre of A.*

Proof. By Lemma 3.5, it's enough to show that $F \cap \partial_H A = \varnothing$.

If not, we could find a path π in F connecting x to $\partial_H A$ never entering twice the same piece of F. We can straighten this to a path, π' in A, with $l(\pi') \leq \nu W(F)$, giving the contradiction that $D(x, A) \leq \nu W(F)$. □

Indeed continuing the same argument, we see easily that $D(G, A) \geq D(x, A) - \nu W(F)$.

Definition. A horizontal surface, F, is said to be *narrow* if $W(F) < W_0$.

Thus the analysis in Section 4.2 tells us that every point of $\Theta(M)$ is contained in a narrow fibre.

Let $D_0 = \nu W_0$.

A particular case of Lemma 3.1 and the subsequent remark is:

Corollary 5.2. *If A a band and S is a narrow horizontal fibre and $x \in S \cap A$ with $D(x, A) \geq D_0$, then $S \cap A$ is a horizontal fibre of A. Moreover, $D(S \cap A, A) > D(x, A) - D_0$.* □

In particular every point of depth at least D_0 in A is contained in a narrow horizontal fibre of A.

Definition. A band B is *narrow* if $W(B) < W_0$.

Recall, from Section 1, the definition of a "collared band", $B \subseteq \hat{B}$, where $\hat{B} = B_- \cup B \cup B_+$ and B_- and B_+ are the top and bottom collars of B. Note that $D(B, \hat{B}) = \min\{H(B_-), H(B_+)\}$. We say that B is *narrow* as a collared band if both B and $\hat{B}$ are narrow. We say that B is *h-collared* if $D(B, \hat{B}) \geq h$.

We will observe that sufficiently long bands will always contain parallel collared bands of bounded width. This will ultimately reduce us to considering only collared bands. One advantage of this is that they satisfy a certain nesting property, stated in Lemma 5.3 below. This nesting property, a priori, only applies to base surfaces. The bands themselves need not be nested. This is a complicating factor, that will need to be addressed later (after the proof of Lemma 5.4.)

Lemma 5.3. *Suppose $h \geq 0$ and A is a band with $H(A) \geq 2h + 4D_0$. Then A contains a narrow band B with $h \leq D(B, A) \leq h + D_0$ and with $H(B) \geq H(A) - 2h - 4D_0$.*

Proof. Let π be a vertical fibre of A with $l(\pi, A) = H(A)$. Let $x_\pm$ be points of $\pi \setminus \bigcup \mathcal{T}_0(A)$ at external distance $h + D_0$ away from $\partial_\pm A$. By Corollary 5.2 there are narrow horizontal fibres, $F_\pm$, of A containing $x_\pm$. As in the proof of Lemma 5.1, we see that $H(\langle F_\pm, \partial_\pm A\rangle) \geq h + D_0 - D_0 = h$ and $H(\langle F_-, F_+\rangle) \geq H(A) - 2(h + D_0) - 2D_0 = h - 4D_0$. We set $B = \langle F_-, F_+\rangle$. □

In particular, if we set $\hat{B} = A$, we get an h-collared band, $(B, \hat{B})$.

Lemma 5.4. *Suppose that B_1, B_2 are narrow primitive $(2D_0)$-collared bands. If $B_1 \cap B_2 \neq \varnothing$, then either $\phi(B_1) \subseteq \phi(B_2)$ or $\phi(B_2) \subseteq \phi(B_1)$.*

Proof. Let $x \in B_1 \cap B_2 \cap \Theta(M)$. By Section 4.2, x lies in a narrow horizontal fibre, S, of M. Now $D(x, \hat{B}_i) \geq D_0$ and so by Corollary 5.2, $F_i = S \cap \hat{B}_i$ is a horizontal fibre of $\hat{B}_i$ containing x. Moreover, $D(F_i, \hat{B}_i) \geq 2D_0 - D_0 = D_0$. Let G be the component of $F_1 \cap F_2$ containing x. Thus, $\phi(G)$ is a component of $\phi(B_1) \cap \phi(B_2)$.

Suppose that $\phi(B_1)$ is not a subset of $\phi(B_2)$, or equivalently that $F_1 \not\subseteq F_2$. There must be a boundary curve, say α, of G contained in the interior of F_1. Thus, $\alpha \subseteq \partial F_2$. Now α is a longitude of some $T \in \mathcal{T}$. Since F_2 is a fibre of $\hat{B}_2$, $T \in \mathcal{T}_\partial(\hat{B}_2) = \mathcal{T}_\partial(B_2)$. Since α lies in the interior of F_1, $T \notin \mathcal{T}_\partial(B_1)$. Moreover, from the last paragraph, we see that $D(\partial T, \hat{B}_1) \geq D_0$.

Now $T \in \mathcal{T}_\partial(\partial_\pm B_2)$. Let $\gamma_\pm \subseteq T \cap \partial_\pm B_2$ be longitudes of ∂T on the same side of T as α, i.e. so that γ_+, γ_- and α all lie in the same component of $\partial_V \hat{B}_2$. Since $D(\gamma_\pm, \hat{B}_1) \geq D_0$, by Lemma 5.1, it follows that there are horizontal subsurfaces, $G_\pm$, of $\partial_\pm$ containing $\gamma_\pm$, so that $\phi(G_+)$ and $\phi(G_-)$ are both components of $\phi(B_1) \cap \phi(\partial_\pm B_2) = \phi(B_1) \cap \phi(B_2)$. Since γ_+ and γ_- are on the same side of T, with respect to B_2, they must map to the same boundary curve of $\phi(B_2)$. In particular, $\phi(G_+) \cap \phi(G_-) \neq \varnothing$, and so $\phi(G_+) = \phi(G_-) = J$, say.

Now if $\phi(B_2) \not\subseteq \phi(B_1)$, there must be some boundary curve, β, of J lying in the interior of $\phi(B_2)$. We have $\beta = \phi(T)$ for some $T \in \mathcal{T}$. Since $\beta \subseteq \partial\phi(G_\pm)$ there must be curves $\delta_\pm \subseteq \partial G_\pm$ which are longitudes in T. Since $\delta_\pm$ are not boundary curves of $\partial_\pm B_2$, It follows that $T \in \mathcal{T}_I(\partial_\pm B_2) = \mathcal{T}_\pm(B_2)$. In particular, $\mathcal{T}_+(B_2) \cap \mathcal{T}_-(B_2) \neq \varnothing$, contradicting the assumption that B_2 is primitive. $\square$

We remark that by the same argument, we can arrive at the same conclusion assuming, for any $k > 0$, that B_1 and B_2 are $(2D_0 + k)$-collared, and that $d(B_1 \cap \Theta(M), B_2 \cap \Theta(M)) \leq k$.

It would be nice if we could go on to conclude that collared bands were nested. However, it is still possible that a horizontal boundary component of the larger band (the one with larger base surface) may cut through the smaller band. This is a phenomenon that will need to be described and dealt with. This is the purpose of the following discussion.

Let B be a band. Note that the horizontal boundary, $\partial_H B = \partial_+ B \cup \partial_- B$ meets $\Theta(M)$ precisely in the relative boundary of $B \cap \Theta(M)$ in $\Theta(M)$. If A is a primitive band, then $A \cap \Theta(M)$ is connected. We see easily that one of $A \subseteq B$, $A \cap B = \varnothing$ or $A \cap \partial_H B \neq 0$ must hold.

Recall that A, B are *nested* if $A \subseteq B$, $B \subseteq A$ or $A \cap B = \varnothing$. Suppose that A, B are nonnested primitive narrow $(2D_0)$-collared bands. Since $A \cap B \neq \varnothing$, applying Lemma 5.4, we have either $\phi(A) \subseteq \phi(B)$ or $\phi(B) \subseteq \phi(A)$. Suppose that $\phi(A) \subseteq \phi(B)$. Since $A \not\subseteq B$, we have $A \cap \partial_H B \neq \varnothing$, so without loss of generality, $A \cap \partial_+ B \neq \varnothing$. Applying Corollary 5.2, we see that $F = \hat{A} \cap \partial_+ B$ is a horizontal subsurface of B that is a fibre of $\hat{A}$. In particular, $\mathcal{T}_\partial(F) = \mathcal{T}_\partial(A) \subseteq \mathcal{T}_\partial(\partial_+ B) \cup \mathcal{T}_I(\partial_+ B) = \mathcal{T}_\partial(B) \cup \mathcal{T}_+(B)$.

Before continuing, we remark that the same argument would apply if we assume that A and B are not k-nested and that A and B are $(2D_0 + k)$-collared. Note that $N(A \cap \Theta(M), k)$ is connected, so in this case $d(A \cap \Theta(M), \partial_H B \cap \Theta(M)) \leq k$, which is sufficient to make the argument work.

Now suppose that $\phi(A) \neq \phi(B)$, so that F is a proper subsurface of $\partial_+ B$. We see that $\mathcal{T}_\partial(A) \setminus \mathcal{T}_\partial(B) \neq \varnothing$. Note that $\mathcal{T}_\partial(A) \setminus \mathcal{T}_\partial(B) \subseteq \mathcal{T}_+(B)$. (We also remark that it follows that $A \cap \partial_- B = \varnothing$, otherwise a similar argument would give $\mathcal{T}_\partial(A) \setminus \mathcal{T}_\partial(B) \subseteq \mathcal{T}_-(B)$, showing that $\mathcal{T}_\partial(B) \cap \mathcal{T}_-(B) \neq \varnothing$, and contradicting the assumption that B is primitive.) One can also see easily that $A \subseteq \hat{B}$ (see Lemma 5.5 below).

Now let B' be the band with $\partial_- B' = \partial_- B$ and with $\partial_+ B'$ the horizontal surface obtained from $\partial_+ B$ by replacing $F \subseteq \partial_+ B$ with the parallel surface $\partial_- A$, pushed downwards slightly so that if becomes disjoint from $\partial_- A$. The remainder of $\partial_+ B$ remains unchanged apart from suitable adjustments of the annuli in the tubes of $\mathcal{T}_\partial(A) \setminus \mathcal{T}_\partial(B)$. We can assume that B' remains narrow. Clearly $\phi(B') = \phi(B)$. In fact:

Lemma 5.5. *We have $A \cap B' = \varnothing$ and $B' \subseteq \hat{B}$. Moreover, $H(B) \leq H(B')$ and $D(B, \hat{B}) \leq D(B', \hat{B})$.*

Proof. That $A \cap B' = \varnothing$ follows easily from the construction. Let $h = D(B, \hat{B})$. Choose some $T \in \mathcal{T}_\partial(A) \cap \mathcal{T}_I(B)$. Now $D(T, \hat{B}) \geq h$. Moreover $T \in \mathcal{T}_\partial(\partial_\pm \hat{A})$ and $\hat{A}$ is a assumed to be narrow. Thus

$$D(\hat{A}, \hat{B}) = D(\partial_H \hat{A}, \hat{B}) \geq h - D_0 \geq D_0.$$

In particular, $\hat{A} \subseteq \hat{B}$ and so $B' \subseteq B$.

Let π be a path in $\hat{B} \setminus \partial_V \hat{B}$ with (close to) minimal external length $l(\pi)$ that connects $\partial_- B = \partial_- B'$ to $\partial_+ B'$. Let x be its endpoint in $\partial_+ B'$. Now if x lies the subsurface we pushed off $\partial_- A$, then π has to cross $A_- = \langle \partial_- A, \partial_- \hat{A} \rangle$. This contributes at least (almost) $2D_0$ to $l(\pi)$, so it would have been quicker simply to follow $\partial_+ \hat{A}$ to T (straightening in $\Theta(M)$) and then go through T to reach the unaltered part of $\partial_+ B'$. In other words, we arrive at a point of $\partial_+ B$, and so $H(B') \geq L(\pi) \geq H(B)$.

The fact that $D(B', \hat{B}) \geq D(B, \hat{B})$ is similar, but even simpler. Note that it would be stupid for a vertical fibre of B'_+ to go all the way through A_+ and A in order to reach $\partial_- A$, when it could just go directly to T. Moreover, the bottom collar, B_- remains unchanged. □

These results show us how to arrange any pair of $(2D_0)$-collared bands to be nested, except possibly if $\phi(A) = \phi(B)$. The construction of the band system will involve choosing at most one band (of almost maximal height) with a given base surface, so that the last situation will not arise.

Recall that we have fixed constants W_0, D_0 and H_3 depending on $\kappa(\Sigma)$ (as described in Section 4). We fix further constants, $H_0 \geq 2D_0$ and $H_1 \geq 0$, and let $H_4 = H_1 + 2H_0 + 4D_0$. We assume that $H_4 \geq H_3$.

The aim is to construct a set, $\mathcal{B}$, of bands satisfying:

(B1) The elements of $\mathcal{B}$ are nested.

(B2) No two elements of $\mathcal{B}$ have the same base surface.

(B3) Each element of $\mathcal{B}$ is a narrow H_0-collared band.

(B4) Each $B \in \mathcal{B}$ has $H(B) \geq H_1$.

(B5) If F is a narrow horizontal surface parallel to $B \in \mathcal{B}$, then either

$$H(\langle F, \partial_+ B\rangle) \leq H_0 + 2D_0 \quad \text{or} \quad H(\langle F, \partial_- B\rangle) \leq H_0 + 2D_0.$$

(B6) If A is any narrow band with $H(A) \geq H_4$, then there is some band in $\mathcal{B}$ with the same base surface.

To construct $\mathcal{B}$, let $\mathcal{F}_0$ be the set of $\Phi \in \mathcal{F}$ for which there is a narrow band, A, with $\phi(A) = \Phi$ and $H(A) \geq H_4$. For convenience, we assume that the maximal height is attained, say by A. Lemma 5.3 then gives us a subband $B \subseteq A$, so that setting $\hat{B} = A$, we get a H_0 collared band, with $H(B) \geq H(A) - 2D_0 - 4D_0 \geq H_1$. Properties (B2)–(B6) are more or less immediate. To obtain nestedness, (B1), we need to carry out the modification procedure described above. We start with bands with base surfaces of minimal complexity, and proceed inductively over complexity. A given band B might meet other bands A with $\phi(A)$ strictly contained in $\phi(B)$. Inductively, the set of all such bands A meeting B is nested. We can thus perform the construction described before Lemma 5.5 to the set of outermost such bands simultaneously (or in any order) to give us a band, B'. We now replace B by B'. After a finite number of such modifications, we arrange that B is nested with all other bands. We do this for all bands with the same complexity, and then move on to bands with the next higher complexity. By the time we reach $\kappa(\Sigma) - 1$, we obtain our nested band system $\mathcal{B}$.

Note that the existence of collars (B3) also implies that $d(\partial_+ B \cap \Theta(M), \partial_- B \cap \Theta(M)) \leq H_0$; thus a band does not approach itself on the outside.

It is possible to refine the procedure above slightly. As we have stated it, if A and B are bands with $\phi(A) \subseteq \phi(B)$, then it is possible for $\partial_- A$ to be very close to $\partial_+ B$. There is a slight modification of the process that will ensure that $d(A \cap \Theta(M), B \cap \Theta(M)) \geq H_2$ for an arbitrarily chosen constant, $H_2 > 0$. To achieve this, we construct our initial bands to be doubly collared. In other words, for each $B \in \mathcal{B}$ is initially contained in two larger bands, $B \subseteq \bar{B} \subseteq \hat{B}$, with $H_2 \leq D(B, \bar{B}) \leq H_2 + D_0$ and $H_1 \leq D(\bar{B}, \hat{B}) \leq H_2 + D_0$. If $A \in \mathcal{B}$ with $\bar{A} \cap \partial_+ B \neq 0$, then we can assume there is a horizontal subsurface, F, of $\partial_+ B$, with $\phi(F) = \phi(A)$.

We modify $\partial_+ B$ replacing F by $\partial_- \bar{A}$. Similarly, if $C \in \mathcal{B}$ with $\bar{C} \cap \partial_+ \bar{B} \neq 0$, then there is a subsurface, G, of $\partial_+ \bar{B}$ with $\phi(G) = \phi(C)$. We modify $\partial_+ \bar{B}$ by replacing G with $\partial_- \bar{C}$. We do this for all such A and C, and proceed inductively for over the complexity of $\phi(B)$. We can do the same thing for the bottom surfaces (swapping $+$ and $-$). Finally we forget about the intermediate bands, $\bar{B}$, and get a system of collared bands as before.

Putting this together with the earlier remarks, and by taking our bands to be at least $(2D_0 + H_2)$-collared, we can assume that the set $\mathcal{B}$ is H_2-nested.

We want to explore properties of $\mathcal{B}$. Most of the work, carried out in Section 6, is to verify property (A3) of Section 1. We begin here with some preliminary discussion of pushing surfaces off bands. As one consequence of this, we will deduce that our bands are unlinked in M. For the remaining discussion of this section, we will not need (B6). We note that properties (B1)–(B5) pass to any subset of $\mathcal{B}$, in particular to the set of outermost bands of $\mathcal{B}$.

We recall the process of pushing fibres. Suppose that S, Z are horizontal fibres, and $F \subseteq S \setminus Z$ an essential subsurface or curve contained in a piece of S. Let S' be the horizontal surface obtained pushing S off Z as described in Section 3.

Now each piece of S' is obtained by gluing together subsets of pieces of S and Z. Some of the subsets of Z may be discs, but there is a combinatorial bound in terms of $\kappa(\Sigma)$ on the number of nondisc components glued together in this way. Thus, $W(S')$ is bounded above by some (linear) function of $W(S)$ and $W(Z)$. The same discussion applies to pushing S off a horizontal surface K parallel to a horizontal subsurface of S. In this case, we get a (linear) bound in terms of $W(S)$ and $W(K)$.

Now, let $\mathcal{B}$ be a collection of bands satisfying (B1)–(B5) above. Let $\mathcal{A} \subseteq \mathcal{B}$ be the subset of outermost bands. Clearly $\bigcup \mathcal{A} = \bigcup \mathcal{B}$.

Suppose that S is a narrow horizontal surface, and $F \subseteq S \cap \Theta(M) \setminus \bigcup \mathcal{B}$ is an essential subsurface or curve. Suppose that $B \in \mathcal{A}$ and $S \cap B \neq \varnothing$. By Corollary 5.2, $G = S \cap \hat{B}$ is a horizontal fibre of $\hat{B}$. If $F \cap G = \varnothing$, then we can replace G is S by $\partial_+ B$, pushed slightly off B. The fibre S remains narrow. After doing this for each such B, we can assume that $F \cap G \neq \varnothing$, and so $F \subseteq G$. In this case we can apply the pushing construction so as to push S first off $K = \partial_+ B$ and then off $K = \partial_- B$. The resulting fibre still contains F, does not meet B, and has width bounded above in terms (depending on $\kappa(\Sigma)$) of $W(S) \leq W_0$ and $W(B) \leq W_0$. We now apply this successively to all such B. Since they are each parallel to a horizontal subsurface of our original fibre, there is a combinatorial bound on the number of such B in terms of $\kappa(\Sigma)$. We thus finally obtain a fibre, $S' \supseteq F$ with $S' \cap \bigcup \mathcal{B} = \varnothing$, and with $W(S')$ bounded above by some constant W_1 depending only on $\kappa(\Sigma)$.

Putting this together with the property in Section 4.2, we obtain, in particular:

Lemma 5.6. *If $T \in \mathcal{T}$ and α is a horizontal longitude of ∂T disjoint from $\bigcup \mathcal{B}$, then α is contained in a horizontal fibre, S, with $W(S) < W_1$ and $S \cap \bigcup \mathcal{B} = \emptyset$.* □

We can apply this to show that the set, $\mathcal{A}$, of outermost bands of $\mathcal{B}$ are unlinked in M. Given any $A \in \mathcal{A}$, let $S(A)$ be any narrow horizontal fibre in M meeting A in a fibre, $F(A)$, of A. Now the collection of bands $\mathcal{A} \setminus \{A\}$ also satisfies (B1)–(B5) above, so applying the construction above, we can push $S(A)$ off each element of $\mathcal{A} \setminus \{A\}$ while keeping $F(A)$ unchanged. We thus obtain fibres $(S'(A))_{A \in \mathcal{A}}$ with $F(A) \subseteq \mathcal{A}$ and $F(A) \cap S'(B) = \emptyset$ for all distinct $A, B \in \mathcal{A}$. Now Lemma 3.2 tells us that the surfaces, $F(A)$, and hence, by definition, the bands A are unlinked. In other words, we have shown:

Lemma 5.7. *A set of outermost bands satisfying (B1)–(B5) is unlinked in M.*

6. Bounding vertical lengths

The main purpose of this section is show that a set of bands satisfying (B1)–(B6) of Section 5 will also satisfy (A3) of Section 2. Having constructed such a set of bands in Section 5, this will prove the main result, namely Theorem 0.

Given $T \in \mathcal{T}$, recall that $L(\partial T, \mathcal{B})$ is defined as the total vertical length of $\partial T \setminus \bigcup \mathcal{B}$. We aim to show:

Proposition 6.1. *There is some L_0 such that for all $T \in \mathcal{T}$, $L(\partial T, \mathcal{B}) \leq L_0$.*

Here, L_0 depends on $\kappa(\Sigma)$ and the choice of H_0 and H_1.

Convention. Throughout this section, we will use the term "band" only to refer to elements of $\mathcal{B}$. Other bands (as we have defined them) will be termed "strips". Unless otherwise stated, each "horizontal surface" will be assumed disjoint from $\bigcup \mathcal{B}$, and any strip will be assumed nested with the elements of $\mathcal{B}$, and not contained in any element of $\mathcal{B}$.

A horizontal surface, F, will be said to be "narrow" if its width, $W(F)$, is less than W_1.

We have thus strengthened the notion of "horizontal surface", but weakened the definition of "narrow" (as given in Sections 2 and 5 respectively). By Lemma 5.6, it remains the case that every point of M lies in a narrow horizontal surface.

To exploit bounded geometry, we will use the following variation of the nerve of a covering. The construction will also be used in Section 7. Recall, from Section 4, the definition of an "r-net", $V \subseteq \Theta(M)$. It will be convenient to construct V as follows. Given $T \in \mathcal{T}$, let $V(\partial T)$ be an r-net in ∂T. The condition on r ensures that $\bigcup_{T \in \mathcal{T}} V(\partial T)$ is r-separated in $\Theta(M)$. We now extend $\bigcup_{T \in \mathcal{T}} V(\partial T)$ to an r-net, V, for $\Theta(M)$.

Let Δ be the graph with vertex set $V(\Delta) = V$ and $x, y \in V$ adjacent if $d(x, y) \leq 3r$ in $\Theta(M)$. The vertices of Δ have bounded degree. Note that if $R \subseteq \Theta(M)$, then $R \subseteq N(P, r)$, where $P = V \cap N(R, r)$.

Given $T \in \mathcal{T}$, let $\Upsilon(\partial T)$ be the complete graph on $V(\partial T) = V \cap \partial T$. Let $\Upsilon = \Delta \cup \bigcup_{T \in \mathcal{T}} \Upsilon(\partial T)$. In other words, Υ has the same vertex set, V, but we have added more edges across the Margulis tubes.

The idea behind this construction is that Υ approximates the geometry of M after each Margulis tube has been shrunk to bounded diameter. Lengths in Υ thus correspond to exterior lengths in M to within linear bounds. Here is a more precise formulation.

If p is a path in Υ, then we obtain a path $\pi = \pi(p)$ in M as follows. Suppose x, y are adjacent vertices of p. If the edge between them lies in Δ, then we connect x to y by a path of length at most $3r$ in $\Theta(M)$. If it lies in $\Upsilon(\partial T)$ for some $T \in \mathcal{T}$, then we connect x to y by any path in the interior of T. (Its homotopy class in T will not be important.)

Conversely, given any path in M, recall that $l(\pi)$ is its exterior length, i.e. the length of $\pi \cap \Theta(M)$. We can find a path $p = p(\pi)$ in Υ, whose combinatorial length is at most $l(\pi)/r$ and for which $\pi(p(\pi))$ remains within a distance $3r$ of π in $\Theta(M)$.

We also recall the straightening process used in Section 5, for example in the proof of Lemma 5.1. If π is a path in M, then we can replace any segment of $\pi \cap \Theta(M)$ by a shortest path with the same endpoints, give us another path π' (not assumed to be homotopic to π). Thus, $l(\pi')$ will be at most the sum of the diameters of the components of $\pi \cap \Theta(M)$. This straightening is necessary because the bounds on width refer only to the ambient diameters in $\Theta(M)$ rather than intrinsic diameters. However, it is a technical point that can be ignored for the purposes of following the overall logic.

Here is a key step in the proof of Proposition 6.1:

Lemma 6.2. *Given $L, W \geq 0$, there is some $E = E(L, W)$ with the following property. Suppose that A is a strip in M with $\phi(A) \neq \Sigma$ and $W(A) \leq W$. Suppose that $L(\partial T \cap A, \mathcal{B}) \leq L$ for all $T \in \mathcal{T}_0(A)$. If $T' \in \mathcal{T}_\partial(A)$, then $L(\partial T' \cap A, \mathcal{B}) \leq E$.*

(Recall that $\mathcal{T}_0(A) = \mathcal{T}_I(A) \cup \mathcal{T}_+(A) \cup \mathcal{T}_-(A)$.)

Recall that our eventual aim is to prove (A3), namely that the vertical length of the boundary of each Margulis tube in the exterior of the bands is bounded. This lemma will deal with the inductive step in the argument. It says that if we know this for the intersections of tubes in $\mathcal{T}_0(A)$, then we know it also for the tubes in $\mathcal{T}_\partial(A)$.

Lemma 6.2 would follow fairly easily if we could bound the total volume of $A \setminus \bigcup \mathcal{B}$, in other words, the number of components of $A \setminus \bigcup \mathcal{B}$, and the volume of each such component. For the latter, by general principles of bounded geometry, it would be enough to bound the diameter of each component. We can deal with these two issues simultaneously by making a combinatorial approximation to the geometry. This uses the graph, Υ, defined above. We can reinterpret the volume in terms of the number of vertices. To bound this, in turn, it is sufficient to bound the diameter of Υ and the degree of its vertices. In the case where the height, $H(A)$, is bounded, the diameter bound follows from the fact that every fibre of M must meet $A_- \cup \pi \cup A_+$, where π is any vertical fibre of A. For the degree bound, we use the bounded geometry of $\Theta(M)$, together with the hypothesis on the Margulis tubes in $\mathcal{T}_0(A)$. If $H(A)$ is very large, on the other hand, by construction of the band system, there will be a band, $B \in \mathcal{B}$, with the same base surface as A. We can then apply the above to the components of $A \setminus B$.

We now give a formal proof.

Proof of Lemma 6.2. First suppose that $H(A)$ is less than some constant H, and give a bound in terms of L, W and H. (Note that we allow the possibility for A be nonprimitive, i.e. $H(A) = 0$.)

Let π be a vertical fibre of $A \setminus \partial_V A$ with $l(\pi) \leq H$. Let $a_\pm$ be its endpoint in $\partial_\pm A$. We can assume that π meets the boundary of each Margulis tube in at most two points.

If $x \in \partial_\pm A$, then we can connect x to $a_\pm$ by a path, π, in $\partial_\pm A$ which only enters Margulis tubes in $\mathcal{T}_0(A)$, and then at most once. We can thus straighten to π to a path π' in $\Theta(M)$, with $l(\pi') \leq \nu W(A) \leq \nu W$, and which only meets boundaries of Margulis tubes in points of $A \setminus \bigcup \mathcal{B}$.

Suppose $y \in A \cap \Theta(M) \setminus \bigcup \mathcal{B}$. By Lemma 5.6, y is contained in a narrow horizontal fibre of M (in the sense above). This fibre must intersect $\pi \cup \partial_H A$ at some point x. As above, y can be connected by a path of exterior length at most νW_1, and entering and leaving Margulis tubes only in points of $A \setminus \bigcup \mathcal{B}$.

We see that any two points, $v, w \in A \setminus \bigcup \mathcal{B}$ can be connected by a path τ in M with $l(\tau) \leq H + 2\nu W + 2\nu W_1$, and if τ meets $T \in \mathcal{T}$, then $T \in \mathcal{T}_0(A)$ and $\tau \cap \partial T \subseteq R(\partial T)$, where $R(\partial T) = (\partial T \cap \pi) \cup (\partial T \cap A \setminus \bigcup \mathcal{B})$. If $v, w \in V$, then we can connect v and w by a path $p = p(\pi)$ in Υ of combinatorial length at most $l(\tau)/r$. Moreover, if x, y are adjacent vertices of p connected by an edge in $\Upsilon(\partial T)$ for some $T \in \mathcal{T}$, then $T \in \mathcal{T}_0(A)$ and x and y lie in $N(R(\partial T), 2r)$. But now $\pi \cap \partial T$ consists of at most two points, and by assumption, the vertical length of $A \cap \partial T \setminus \bigcup \mathcal{B}$ is at most L. It follows that there is a universal bound, in terms of L, for the number of possible x and y, and hence on the number of possible edges along which p can cross $\Upsilon(\partial T)$.

Given the bound on the degrees of vertices in Δ and on the length of p, we see that there is a bound on the number of possibilities for such a path p, in terms of L, W and H. This bounds the cardinality of V in terms of L, W, H. Indeed (given the bound on the degrees of vertices in Δ), we get a bound on the cardinality of $P = V \cap N(A \setminus \bigcup \mathcal{B}, 2r)$.

But now, since V is an r-net in $\Theta(M)$, we have $A \cap \Theta(M) \subseteq N(P, r)$. In particular, if $T \in \mathcal{T}_\partial(A)$, then $\partial T \cap A \setminus \bigcup \mathcal{B} \subseteq N(P, r)$. But the intersection of ∂T with $N(P, r)$ has bounded area for all $x \in \Theta(M)$. This places a bound on the area, and hence vertical length of $\partial T \cap A \setminus \mathcal{B}$ in terms of L, W and H as claimed.

We finally need to remove the dependence on the height of A.

First, suppose there is no band of $\mathcal{B}$ with base surface $\phi(A)$. By property (B6), this means that any strip, B, with $\phi(B) = \phi(A)$ and with $W(B) \le W_0$ must have $H(B) \le H_4 = H_1 + 2H_0 + 4D_0$. Applying Lemma 5.3 (with $h = 0$) we see that $H(A) \le H_4 + 4D_0$. Thus, we can apply the preceding result with $H = H_4 + 4D_0$.

Secondly, suppose there is some $B \in \mathcal{B}$ with $\phi(A) = \phi(B)$. There are two subcases. Either $B \subseteq A$ or $A \cap B = \varnothing$.

Suppose first that $B \subseteq A$. Now, B has two collars in A, namely $A_+ = \langle \partial_+ B, \partial_+ A \rangle$ and $A_- = \langle \partial_+ B, \partial_- A \rangle$. Consider A_+. By hypothesis, $W(\partial_+ A_+) = W(\partial_+ A) \le W$, and $W(\partial_- A_+) = W(\partial_+ B) \le W_0$. We can assume that $W \ge W_0$, so $W(A_+) \le W$. We also have $H(A_+) \le H_0 + 4D_0$, otherwise, as in Lemma 5.3, we could find a horizontal fibre, F, in A_+ with $D(F, \partial_+ B) \ge H_0 + 3D_0$ and $W(F) \le W_0$. In particular, $\langle F, A_+ \rangle$ would be narrow and of height greater than $H_0 + 2D_0$, in contradiction to (B5). (Note that we cannot apply (B5) directly to $F = \partial_+ A$, since the bound on its width might not be sufficient — W may be bigger than W_0.) We can now see that the hypotheses of the lemma are satisfied by the band A_+, since if $T \in \mathcal{T}_0(A_+) \subseteq \mathcal{T}_0(A)$, then $L(\partial T \cap A_+, \mathcal{B}) \le L(\partial T \cap A, \mathcal{B}) \le L$. The bounded height case of the lemma now shows that if $T' \in \mathcal{T}_\partial(A) = \mathcal{T}_\partial(A_+)$, then $L(\partial T' \cap A_+, \mathcal{B})$ is bounded. Similarly we see that $L(\partial T' \cap A_-, \mathcal{B})$ is bounded. But $L(\partial T' \cap A, \mathcal{B}) = L(\partial T' \cap A_+, \mathcal{B}) + L(\partial T' \cap A_-, \mathcal{B})$, and the result follows in this case.

The remaining case is when $A \cap B = \varnothing$. But now a similar argument, using Lemma 5.3 and (B5) shows that $H(A)$ is bounded, and we are reduced to the earlier case. □

We return to the pushing process. We say that a strip, C, is *full* if $\phi(C) = \Sigma$. Suppose that C is a full strip, that $S \subseteq M$ is a fibre, and that $F \subseteq S \cap \Theta(M)$ is an essential curve or surface. Pushing S successively off $\partial_+ C$ and $\partial_- C$, we obtain another fibre, $S' \supseteq F$, with $W(S')$ bounded above in terms of $W(S)$ and $W(C)$. (As usual, S, $\partial_H C$ and S' are all assumed disjoint from $\bigcup \mathcal{B}$.)

Applying Lemma 5.6, we obtain:

Lemma 6.3. *There is a nondecreasing function, $f : [0, \infty) \to [0, \infty)$ such that if C is a horizontal strip, $T \in \mathcal{T}$ and α is a horizontal fibre of ∂T contained in $A \setminus \bigcup \mathcal{B}$, then α is contained in a horizontal fibre, S of C with $W(S) \le f(W(C))$.* □

Note that S divides C to two full substrips, each of width at most $f(W(C))$.

Given $n \geq 1$, define W_n inductively by $W_{n+1} = f(W_n)$, starting with W_1, the constant of Lemma 5.6.

The following lemma represents the core of the argument. We deal inductively with tubes lying in bigger and bigger strips. To give the idea, suppose, for example, we have some $T \in \mathcal{T}_I(A)$, lying in the interior of a strip, A, with $L(\partial T, \mathcal{B})$ very large. We can find two fibres, S and S' cutting through T, so that on one side, they bound an annulus, $\Omega \subseteq \partial T$, with $L(\Omega, \mathcal{B})$ also very large. Now Ω is the vertical boundary component of a piece, A', the full strip, $\langle S, S' \rangle$, so that $A' \subseteq A$ has smaller complexity. Using induction and Lemma 6.2, we then bound $L(\Omega, \mathcal{B})$, which would give a contradiction. This argument therefore bounds $L(\partial T, \mathcal{B})$. Of course, there are also other cases to be considered. To make the induction hypothesis work smoothly, we shall phrase everything in terms of pieces of full strips. Here is a precise statement:

Lemma 6.4. *Suppose $\kappa \in \{1, \ldots, \kappa(\Sigma)\}$. Suppose that C is a full strip with $W(C) \leq W_{2\kappa(\Sigma)-2\kappa}$, and suppose that A is a piece of C with $\kappa(\phi(A)) \leq \kappa$. Then there is some L_κ such that for all $T \in \mathcal{T}_0(A)$ we have $L(\partial T \cap A, \mathcal{B}) \leq L_\kappa$.*

Here L_κ depends only on κ and $\kappa(\Sigma)$. In the case where $\kappa = \kappa(\Sigma)$, we interpret the statement by setting $A = C = M$, and the conclusion means that $L(\partial T, \mathcal{B}) \leq L_{\kappa(\Sigma)}$ for all $T \in \mathcal{T}$. This will therefore imply Proposition 6.1 on setting $L_0 = L_{\kappa(\Sigma)}$.

Proof of Lemma 6.4. The proof will be by induction on κ. First note that the case $\kappa = 0$ is vacuously true, since $\phi(A)$ is then a 3-holed sphere and so $\mathcal{T}_0(A) = \varnothing$.

Now suppose that we have verified the statement for some $\kappa < \kappa(\Sigma)$. If $\kappa < \kappa(\Sigma) - 1$, let A, C be as in the hypotheses, with $\kappa(\phi(A)) = \kappa + 1$, so that $W(C) \leq W_{2\kappa(\Sigma)-2\kappa-2}$. (The case where $\kappa = \kappa(\Sigma) - 1$ will be commented upon at the end.) Let $T \in \mathcal{T}_0(A)$. We want to bound $l = L(\partial T \cap A, \mathcal{B})$.

Suppose first that $T \in \mathcal{T}_I(A)$. Choose any horizontal longitude, α, of ∂T. By Lemma 6.3, there is a horizontal fibre $S \subseteq C$, containing α with $W(S) \leq f(W(C)) \leq f(W_{2\kappa(\Sigma)-2\kappa-2}) = W_{2\kappa(\Sigma)-2\kappa-1}$. Let β be the other intersection of S with ∂T. This is another horizontal longitude of ∂T. Thus, α and β together bound an annulus, $\Omega \subseteq \partial T$, with $L(\Omega, \mathcal{B}) \geq l/2$.

Let $\alpha' \subseteq \partial T \setminus \bigcup \mathcal{B}$ be the horizontal longitude that cuts Ω into two annuli, each having equal vertical length in the complement of $\bigcup \mathcal{B}$ (Figure 6). This vertical length must be at least $l/4$. As before, α' lies in some fibre, $S' \subseteq C$, disjoint from S, with $W(S') \leq f(W(S)) \leq W_{2\kappa(\Sigma)-2\kappa}$. Let β' be the other intersection of S' with ∂T. Since $S \cap S' = \varnothing$, we see that $\beta' \subseteq \Omega$. Swapping α with β if necessary, we can assume that β' does not lie in the annulus, $\Omega' \subseteq \Omega$, bounded by α and α'. Now S and S' bound a strip, $C' \subseteq C$, with $W(C') \leq W_{2\kappa(\Sigma)-2\kappa}$. Also $T \in \mathcal{T}_+(C') \cap \mathcal{T}_-(C')$ and $\Sigma \subseteq \partial T \cap C'$. Thus, Ω' is a vertical boundary component of some piece, A',

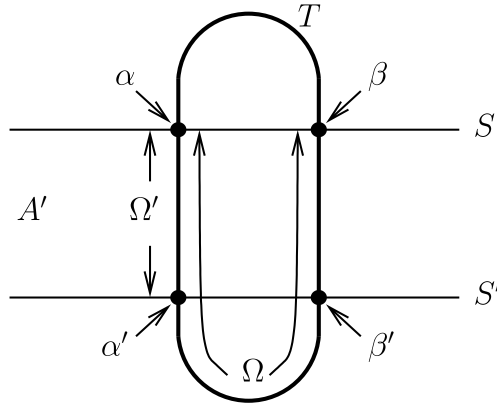


Figure 6

of C'. Thus, $W(A') \le W(C') \le W_{2\kappa(\Sigma)-2\kappa}$. Since $\partial_H A' \cap \partial_H C = \varnothing$, we see that $A' \subseteq A$, and so $\phi(A') \subseteq \phi(A)$. Moreover, since $T \notin \mathcal{T}_I(A')$, $\phi(A') \ne \phi(A)$ and so $\kappa(\phi(A')) < \kappa(\phi(A))$. Thus $\kappa(\phi(A')) \le \kappa$. Now the induction hypothesis tells us that $L(\partial T' \cap A', \mathcal{B}) \le L_\kappa$ for all $T' \in \mathcal{T}_0(A')$. Thus, Lemma 6.2 tells us that $L(\Omega', \mathcal{B}) \le E(L_\kappa, W_{2\kappa(\Sigma)-2\kappa})$, and so $l \le 4L(\Omega', \mathcal{B})$ is bounded as required.

We next consider the case where $T \in \mathcal{T}_\pm(A)$. Without loss of generality, $T \in \mathcal{T}_+(A)$. The discussion only differs from the above in the choice of α and β.

Let $l = L(\partial T \cap A, \mathcal{B})$ as before. Let α' divide $\partial T \cap A$ into two annuli, each of vertical length $l/2$ in the exterior of $\bigcup \mathcal{B}$. Let $S \subseteq C$ be a horizontal fibre containing α with $W(S) \le W_{2\kappa(\Sigma)-2\kappa-1}$, and let β be the other intersection of S with ∂T. Let $\Omega \subseteq \partial T \cap A$ be the annulus bounded by α and $\partial T \cap \partial_+ A$ not containing β. Let C' be the strip bounded by S and $\partial_+ A$, and let A' be the piece of C' containing Ω. Thus Ω is a vertical boundary component of A' (Figure 7). As before, $\kappa(\phi(A')) \le \kappa$ and we get a bound on $L(\Omega, \mathcal{B})$ and hence on l as required.

This proves the induction step when $\kappa < \kappa(\Sigma) - 1$. We can define $L_{\kappa+1}$ in terms of the bounds we have obtained for l.

Finally, we should comment briefly on the final step of the induction, namely when $\kappa = \kappa(\Sigma) - 1$. In this case, we deal with an arbitrary $T \in \mathcal{T}$ in the same way as we did with $T \in \mathcal{T}_I(A)$ above. We obtain two disjoint fibres, S and S', with $W(S) \le W_1$ (by Lemma 5.6) and with $W(S') \le f(W_1) \le W_2$. Thus, $W(C) \le W_2 = W_{2\kappa(\Sigma)-2\kappa}$, and we proceed as before. □

Proof of Proposition 6.1. This is just Lemma 6.4, interpreted for $\kappa = \kappa(\Sigma)$ and setting $L_0 = L_{\kappa(\Sigma)}$. □

Proof of Theorem 0. Let $\mathcal{B}$ be the band system constructed in Section 5, and let $\mathcal{A} \subseteq \mathcal{B}$ be the set of outermost bands. Properties (A1), (A2), (A4), (A5) and (A6) are immediate from the construction, and property (A8) follows directly from

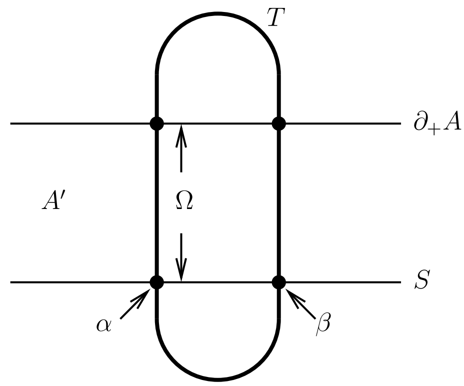


Figure 7

these. Property (A7) follows by Lemma 4.5, and (A9) by Lemma 5.7. Finally (A3) is Proposition 6.1. □

We note that we also have the following relative version, (A3′), for the intrinsic geometry of the band. Again, we suppose that $\mathcal{B}$ satisfies (B1)–(B6). Given $B \in \mathcal{B}$, let $\mathcal{B}(B)$ be the set of bands of $\mathcal{B}$ strictly contained in B. Given $T \in \mathcal{T}_0(B)$, write $L(\partial T \cap B, \mathcal{B}(B))$ for the total vertical length of $\partial T \cap B \setminus \bigcup \mathcal{B}(B)$.

Proposition 6.5. *There is some L_0 such that if $B \in \mathcal{B}$ and $T \in \mathcal{T}_0(B)$, then $L(\partial T \cap B, \mathcal{B}(B)) \leq L_0$.*

Proof. The proof is essentially the same as that of Proposition 6.1. In this case a "horizontal surface" is assumed to be disjoint from $\partial_H B$ and $\bigcup \mathcal{B}(B)$. Only tori in $\mathcal{T}_0(B)$ and bands in $\mathcal{B}(B)$ are relevant to the discussion. □

As mentioned in Section 5, we can also assume (A1′), namely that the bands in $\mathcal{B}$ are H_2-nested.

7. Volume growth

The aim of this section is to prove:

Theorem 7.1. *There is a sequence, $(f_\nu)_{\nu \in \mathbb{N}}$ of polynomials, with f_ν of degree ν, with the following property. Suppose that M is a complete hyperbolic 3-manifold admitting a type-preserving homotopy equivalence to a compact orientable surface Σ, with $\nu(\Sigma) = \nu$. Let $\Theta(M)$ be the thick part of M and* $\mathrm{core}(M)$ *the convex core of M. Suppose that $x \in \mathrm{core}(M) \cap \Theta(M)$ and that $N(x, t)$ is the ball of radius t about x in $\Theta(M)$ for any $t \geq 0$. Then the volume of* $\mathrm{core}(M) \cap N(x, t)$ *is at most $f_\nu(t)$.*

Recall that $\nu(\Sigma)$ is minus the Euler characteristic of Σ. The sequence $(f_\nu)_\nu$ depends only on the choice of Margulis constant. The "type-preserving" condition means that each boundary curve of Σ corresponds to a parabolic cusp of M. The "thick part", $\Theta(M)$, of M consists of M with the interior of the Margulis tubes and Margulis cusps removed. The t-ball, $N(x,t)$, is taken with respect to the induced path metric.

The existence of such a polynomial bound was conjectured by McMullen and proven in [Brock et al. 2004].

The idea of the argument is as follows. Given $B \in \mathcal{B}$, we write $\nu(B) = \nu(\phi(B))$. If $\nu(B) = 1$, the boundaries of the Margulis tubes it contains all have bounded vertical length by (A7) (see [Minsky 1999]) and we see that B has linear growth. We then proceed inductively. For a general band, B, (or M itself) only linearly many outermost subbands $C \subseteq B$ with $\nu(C) < \nu(B)$ are reached in a given time, and by induction, each of these has growth at most polynomial of degree less than $\nu(B)$. Thus the growth rate of B is at most polynomial of degree $\nu(B)$.

There is a subtle issue involved in obtaining the degree, $\nu(\Sigma)$. If one proceeded simply by induction on complexity as previously defined, we would end up with a polynomial of degree $\kappa(\Sigma)$. The refinement arises from the observation that a band, A, may contain a subband, B, whose base surface, $\phi(B)$ is obtained from $\phi(A)$ by removing some set of annuli, so that $\nu(B) = \nu(A)$ (whereas $\kappa(B) < \kappa(A)$). In such a case, $B \cap \Theta(M)$ disconnects $A \cap \Theta(M)$ — a fact that allows us to discount bands of this sort from the discussion. This will be the purpose of Lemma 7.7 below.

To make the argument more precise, it will be convenient to reformulate it in combinatorial terms. We will construct a graph, Π, and a uniform quasi-isometry, $\theta : \Pi \to \Theta(M)$, where Π has growth bounded by a uniform polynomial of degree at most $\nu(\Sigma)$. Here, and in what follows, "uniform" is interpreted to mean dependence only on $\nu(\Sigma)$ and on the Margulis constant defining $\Theta(M)$.

First, we make some general remarks.

Let Π be a graph (not necessarily connected) and let $P \subseteq \Pi$ be a full subgraph (that is, a maximal subgraph with given vertex set). We write Π/P for the quotient graph obtained by collapsing each component of P to a single vertex. (Thus, $\Pi \setminus P$ injects into Π/P.) If $Q \subseteq P$ is full, then

$$\Pi/P = (\Pi/Q)/(P/Q).$$

Also, if $\Pi' \subseteq \Pi$ is any subgraph, then $\Pi' \cap P$ is full in Π', and we write Π'/P for $\Pi'/(\Pi' \cap P)$ viewed as a subgraph of Π/P.

Definition. If Π is a graph and f is a nondecreasing function, we say that Π is $O(f)$ if for all $x \in V(\Pi)$ and all $n \geq 0$, the number of edges in the combinatorial n-ball about x is at most $f(n)/2$. (Note that the degree of Π is bounded above by $f(1)/2$.)

For us, this a convenient way of bounding volume growth in view of the following easily verified lemma.

Lemma 7.2. *Suppose* Π *is a graph and* $P \subseteq \Pi$ *is a full subgraph. If* P *is* $O(f)$ *and* Π/P *is* $O(g)$, *then* Π *is* $O(fg)$. $\square$

Thus if $\varnothing = \Pi_0 \subseteq \Pi_1 \subseteq \cdots \subseteq \Pi_n = \Pi$ is an increasing sequence of full subgraphs and Π_i/Π_{i-1} is $O(f_i)$, then Π is $O(f_1 f_2 \cdots f_n)$.

A subset, Q, of a graph Π is said to be *k-quasidense* in Π if Π is the k-neighbourhood of Q. The *degree* of a graph Π the maximal degree of its vertices. The following is a simple observation.

Lemma 7.3. *Given* $k_1, k_2 \in \mathbb{N}$ *there is a linear function* f *such that if* Π *is a graph of degree at most* k_1 *containing* k_2*-quasidense geodesic*, *then* Π *is* $O(f)$. $\square$

Now let M be a manifold as in the hypotheses of Theorem 7.1. It will be convenient to assume that Σ is closed and that M is doubly degenerate so that $\mathrm{core}(M) = M$. The general case will follow by simple reinterpretation of the arguments.

We will use various graphs that approximate the geometry of M. As before, Δ approximates the thick part, $\Theta(M)$, and Υ approximates the thick part (or M itself) after each Margulis tube has been collapsed to bounded diameter. (These graphs have already been described in Section 6.) These constructions make no reference to our band system $\mathcal{B}$ (other than assuming their vertex sets to be in general position with respect to $\mathcal{B}$). For purely technical reasons, we will introduce another graph, Π, obtained by adding some extra edges to Δ, depending on $\mathcal{B}$. The graphs, Π, and $\Pi \cup \Upsilon$, can also be viewed as approximating $\Theta(M)$, and $\Theta(M)$ with collapsed tubes, respectively. To each band, $B \in \mathcal{B}$, we will associate full subgraphs, $\Delta(B)$ and $\Pi(B)$ of Δ and Π. The purpose of introducing Π is that $\Pi(B)$ will be nicely embedded in Π, whereas it is difficult to ensure that $\Delta(B)$ is nicely embedded in Δ (since our control over the local geometry of $\partial_H B$ is rather weak). For the purposes of understanding the overall logic, one could simply imagine each band of $\mathcal{B}$ to be nicely embedded locally, and just pretend that Δ and Π are identical. We now proceed to a more formal argument.

Let $\mathcal{B}$ be a nested system of bands satisfying (A2)–(A9) and (A1$'$) and (A3$'$) of Section 2.

As in Section 4 we fix some uniform $r > 0$ suitably small in relation to the Margulis constant, as well as the constants featuring in the properties of $\mathcal{B}$. We construct an r-net, V, for $\Theta(M)$ as in Section 6, as follows. First we choose an r-net for ∂T for each $T \in \mathcal{T}$, and then extend $\bigcup_{T \in \mathcal{T}} V(\partial T)$ to an r-net, V, for $\Theta(M)$. We can assume that $V \cap \partial_H B = \varnothing$ for all $B \in \mathcal{B}$.

Let Δ be the graph with vertex set $V(\Delta) = V$ and with $x, y \in V$ adjacent if $d(x, y) \leq 3r$. We construct a map $\theta : \Delta \to \Theta(M)$ as the identity on V and mapping each edge to a (in fact, the) shortest path between its endpoints in $\Theta(M)$. Thus θ is a uniform quasi-isometry.

Given $Q \subseteq M$, write $\Delta(Q)$ for the full subgraph of Δ with vertex set $V \cap Q$. Note that $\bigcup_{T\in\mathcal{T}} \Delta(\partial T)$ is a full subgraph of Δ, and that $\theta(\Delta(\partial T)) \subseteq \partial T$.

Given $B \in \mathcal{B}$, let $E_H(B)$ be the set of edges of Δ with exactly one endpoint in B. Write $V_H(B)$ for those vertices of $\bigcup E_H(B)$ which lie in B. If $e \in E_H(B)$, then $\theta(e)$ crosses $\partial_H B$ (an odd number of times). We can thus partition $E_H(B)$ as $E_+(B) \sqcup E_-(B)$ depending on whether $\theta(e)$ crosses $\partial_+ B$ or $\partial_- B$. We similarly partition $V_H(B)$ as $V_+(B) \cup V_-(B)$.

Given $A \in \mathcal{B}$, let $\mathcal{B}(A) = \{B \in \mathcal{B} \mid B \subseteq A,\ B \neq A\}$, and write $U = \bigcup \mathcal{B}(A) \subseteq B$. Thus $\Delta(U) = \bigsqcup_{B\in\mathcal{B}(A)} \Delta(B)$.

Suppose $T \in \mathcal{T}_0(A)$. By (A8) at most N_0 elements of $\mathcal{B}(A)$ meet ∂T, and by (A3′), $\partial T \setminus U$ has vertical length at most L_0. It follows easily that:

Lemma 7.4. *The quotient graph, $\Delta(\partial T \cap A)/\Delta(\partial T \cap U)$ had uniformly bounded diameter.* ☐

Now $V_H(A) \cap U = \varnothing$ and so we can regard $V_H(A)$ as a subset of $\Delta(A)/\Delta(U)$. Moreover, we can connect $V_+(A)$ to $V_-(A)$ by a path q in $\Delta(A)$ (obtained by approximating any vertical fibre of A by a path in the image of θ). This gives a path $q/\Delta(U)$ from $V_+(A)$ to $V_-(A)$ in $\Delta(A)/\Delta(U)$. Indeed any such path $p \subseteq \Delta(A)/\Delta(U)$ has this form: if p passes through the vertex obtained by collapsing some $\Delta(B) \subseteq \Delta(A)$ we can lift this vertex to a path in $\Delta(B) \subseteq \Delta(A)$ connecting the two incident edges of q.

Recall, from Section 6, that $\Upsilon(\partial T)$ is the complete graph on $V \cap \partial T$, and $\Upsilon = \Delta \cup \bigcup_{T\in\mathcal{T}} \Upsilon(\partial T)$. Given $Q \subseteq M$, write $\Upsilon(Q)$ for the full subgraph of Υ on $V \cap Q$.

Now let q be a path in $\Delta(A)$ connecting $V_+(A)$ to $V_-(A)$. The endpoints of $\theta(q) \subseteq \Theta(M)$ lie within distance $3r$ of $\partial_\pm A \cap \Theta(M)$. It is possible that $\theta(q)$ may cross $\partial_H A$, but by taking a subpath and/or adding short paths to the endpoints, we get a path $\pi \subseteq B \cap \Theta(M)$ connecting $\partial_+ B$ to $\partial_- B$.

Any point $x \in V \cap B$ lies in a horizontal fibre, S, of M with $W(S) \leq W_0$. Clearly $S \cap (\pi \cup \partial_H A) \neq \varnothing$, and so we get a path, s, of bounded length connecting x to $q \cup V_H(A)$ in $\Upsilon(A)$. This path may cross certain graphs $\Upsilon(\partial T)$. However, we can apply Lemma 7.4 to get around these in $\Delta(A \cap \partial T)/\Delta(U \cap \partial T) \subseteq \Delta(A)/\Delta(U)$, adding a bounded amount to the length of $s/\Delta(U)$. Thus x lies a bounded distance from $(q/\Delta(U)) \cup V_H(A)$ in $\Delta(A)/\Delta(U)$. As observed above, any path p from $V_+(A)$ to $V_-(A)$ in $\Delta(A)/\Delta(U)$ has the form $q/\Pi(U)$. We conclude:

Lemma 7.5. *If p is any path from $V_+(A)$ to $V_-(A)$ in $\Delta(A)/\Delta(U)$, then $p \cup V_H(A)$ is uniformly quasidense in $\Delta(A)/\Delta(U)$.* ☐

We would like to say that p is itself quasidense. However there is the technical irritation that the boundary of A may be rather wriggly. We can get around this by adding some extra edges to Δ so as to reduce the diameter of $V_\pm(A)$. This will give us our graph, Π, referred to earlier.

Suppose $B \subseteq \mathcal{B}$, and that F is a piece of $\partial_\pm B$. Let $E(F) \subseteq E_\pm(B)$ be the set of $e \in E_\pm(B)$ such that $\theta(e)$ crosses F. Since $W(\partial_\pm B) = W(B)$ is bounded, so is the diameter of F in $\Theta(M)$, and it follows that $E(F)$ is of bounded diameter in Δ. We extend $E(F)$ to a complete bipartite graph by connecting each vertex of $V \cap B \cap \bigcup E(F)$ to each vertex of $V \cap E(F) \setminus B$. Note that $E_H(B)$ is a disjoint union of such sets $E(F)$. We perform this construction for all such F and all $B \in \mathcal{B}$. This gives us a graph $\Pi \supseteq \Delta$ with the same vertex set V. Moreover (since $W(B)$ is bounded), we can extend θ to a uniform quasi-isometry $\theta : \Pi \to \Theta(M)$. Bounded geometry tells us that Π has uniformly bounded degree. The earlier discussion of Δ applies equally well to Π. In particular, given $Q \subseteq M$, we write $\Pi(Q)$ for the full subgraph of Π on $V \cap Q$. Also we have a graph $\Pi \cup \Upsilon$ on the vertex set V. This time, we see that if $A \in \mathcal{B}$, then $V_\pm(A)$ has bounded diameter in $\Pi \cup \Upsilon$, and so applying Lemma 7.4 as before, we see that it has bounded diameter in $\Pi(A)/\Pi(U)$. Now $\Delta(A)/\Delta(U)$ is a subgraph of $\Pi(A)/\Pi(U)$ with the same vertex set, so putting this together with Lemma 7.5, we deduce:

Lemma 7.6. *Any path connecting $V_+(A)$ to $V_-(A)$ in $\Pi(A)/\Pi(U)$ is uniformly quasidense in $\Pi(A)/\Pi(U)$.* □

This observation is sufficient to tell us that $\Pi(A)/\Pi(U)$ has linear growth (compare Lemma 7.8 below). This, in turn, is enough to give us polynomial growth of Π and hence of $\Theta(M)$ (compare Lemma 7.9). However, to obtain a polynomial of degree $\nu(\Sigma)$, we need to refine this as follows.

Suppose $B \in \mathcal{B}(A)$ with $\nu(B) = \nu(A)$. Now $\partial_\pm B$ can be extended to a horizontal fibre of A by adding a number of annuli in Margulis tubes (in $\mathcal{T}_\partial(B) \setminus \mathcal{T}_\partial(A)$). This follows from the condition that $\nu(B) = \nu(A)$. (Indeed we can extend B to a nonprimitive band $C \subseteq A$ with $\phi(C) = \phi(A)$ by adding some subsolid tori in Margulis tubes.) It follows that $\partial_+ B \cap \Theta(M)$ and $\partial_- B \cap \Theta(M)$ both separate $\partial_+ A \cap \Theta(M)$ from $\partial_- A \cap \Theta(M)$ in $A \cap \Theta(M)$. In other words, any path from $\partial_+ A$ to $\partial_- A$ in $A \cap \Theta(M)$ must pass through B. Interpreting this in terms of the graph Π, we see that any path from $V_+(A)$ to $V_-(A)$ in $\Pi(A)$ contains a subpath connecting $V_+(B)$ to $V_-(B)$ in $\Pi(B)$. It is possible that B may itself contain other subbands of this type, so we will need to give an inductive argument.

Now let $\mathcal{B}_0(A) = \{B \in \mathcal{B}(A) \mid \nu(B) < \nu(A)\}$ and write $U_0 = \bigcup \mathcal{B}_0$. We refine Lemma 7.6 as follows: (If one does not care about the degree of the polynomial, one can go straight to Lemma 7.8, replacing $\mathcal{B}_0$ by $\mathcal{B}$, U_0 by U and ν by κ.)

Lemma 7.7. *Any path connecting $V_+(A)$ to $V_-(A)$ in $\Pi(A)/\Pi(U_0)$ is uniformly quasidense in $\Pi(A)/\Pi(U_0)$.*

Proof. There is a uniform combinatorial bound on the length of a strictly increasing sequence of bands, $B_1 \subset B_2 \subset \cdots \subset B_n = A$ with $B_i \in \mathcal{B}$ and $\nu(B_1) = \nu(A)$. We prove Lemma 7.7 by induction in the maximal such length, $n = n(A)$.

If $n = 1$, then $\mathcal{B}_0(A) = \mathcal{B}(A)$, so Lemma 7.7 reduces to Lemma 7.6.

Suppose $n(A) = n$, and we have verified the lemma for $n - 1$. Let p be any path in $\Pi(A)/\Pi(U_0)$ from $V_+(A)$ to $V_-(A)$. This has the form $q/\Pi(U_0)$, where q connects $V_+(A)$ to $V_-(A)$ in $\Pi(A)$. Let $\mathcal{B}_1(A)$ be the set of bands $C \in \mathcal{B}(A)$ that are outermost in $\mathcal{B}(A)$ and satisfy $\nu(C) = \nu(A)$. Thus

$$U = \bigcup \mathcal{B}(A) = \bigcup (\mathcal{B}_0(A) \cup \mathcal{B}_1(A)) = U_0 \cup U_1,$$

where $U_1 = \bigcup \mathcal{B}_1(A)$.

Suppose $C \in \mathcal{B}_1(A)$. Then $n(C) = n - 1$ and $\mathcal{B}_0(C) = \mathcal{B}_0(A) \cap \mathcal{B}(C)$. Thus $\Pi(C)/\Pi(U_0) = \Pi(C)/\Pi(\bigcup \mathcal{B}_0(C))$. Now q contains a subpath, q_C, connecting $V_+(C)$ to $V_-(C)$ in $\Pi(C)$. By the induction hypothesis, q_C is uniformly quasidense in $\Pi(C)/\Pi(U_0)$.

By Lemma 7.6, $q/\Pi(U)$ is uniformly quasidense in $\Pi(A)/\Pi(U)$. Thus, if $x \in V \cap A$, then x can be connected to q by a path s in $\Pi(A)$ with $s/\Pi(U)$ of bounded length. If $q \cap \Pi(U_1) = \varnothing$. then $s/\Pi(U) = s/\Pi(U_0)$ and we are happy. If not, then s enters some $C \in \mathcal{B}_1(A)$ for the first time at some $y \in \Pi(C)$. From the previous paragraph, we see that there is a path, t, from y to q in $\Pi(C)$ with $t/\Pi(U_0)$ of bounded length. By joining together $s/\Pi(U_0)$ and $t/\Pi(U_0)$ we see that x is a bounded distance from $q/\Pi(U_0)$ in $\Pi(A)/\Pi(U_0)$, and the lemma follows by induction. $\square$

Another point to note is that since $W(B)$ is bounded for all $B \in \mathcal{B}$, there is a bound on the number of edges e of Π such that $\theta(e)$ crosses $\partial_H B$. Thus there is a bound on the number of edges of Π with exactly one endpoint in $\Pi(B)$, and hence on the degree of $\Pi/\Pi(\bigcup \mathcal{B}')$ for any subset $\mathcal{B}'$ of $\mathcal{B}$. In particular, the degree of $\Pi(A)/\Pi(U_0)$ is uniformly bounded.

Putting this observation together with Lemma 7.3 and Lemma 7.7, taking any shortest path from $V_+(A)$ to $V_-(A)$ in $\Pi(A)/\Pi(U)$, we conclude:

Lemma 7.8. *There is a uniform linear function, f, such that for all $A \in \mathcal{B}$, the quotient $\Pi(A)/\Pi(U_0)$ is $O(f)$, where $U_0 = \bigcup \mathcal{B}_0(A)$.* $\square$

(Here f, may depend on $\nu(\Sigma)$.)

Now exactly the same argument applies to M itself, taking a bi-infinite geodesic in $\Pi/\Pi(U_0)$, where $U_0 = \bigcup \mathcal{B}_0$, and $\mathcal{B}_0 = \{B \in \mathcal{B} \mid \nu(B) < \nu(\Sigma)\}$. Thus, $\Pi/\Pi(U_0)$ is also $O(f)$.

Now, given $n \in \{1, 2, \dots, \nu(\Sigma) - 1\}$, let $\mathcal{B}_n = \{B \in \mathcal{B} \mid \nu(B) = n\}$. Let $\mathcal{C}_n \subseteq \mathcal{B}_n$ be the set of bands of $\mathcal{B}_n$ that are outermost, and let $\mathcal{C} = \bigcup_{n=1}^{\nu(\Sigma)-1} \mathcal{C}_n$. Thus if $A, B \in \mathcal{C}$ with B strictly included in A, then $\nu(B) < \nu(A)$. If $A \in \mathcal{C}$ then $\mathcal{B}_0(A) = \mathcal{B}(A) \cap \mathcal{C}$.

Given n, let $U_n = \bigcup \mathcal{C}_n = \bigcup \mathcal{B}_n$, and let $\Pi_n = \Pi(U_n)$. Each component of Π_n has the form $\Pi(A)$ for some $A \in \mathcal{C}_n$. Each component of Π_{n-1} inside Π_n has the form $\Pi(B)$ for some $B \in \mathcal{B}(A) \cap \mathcal{C} = \mathcal{B}_0(A)$. Thus $\Pi_{n-1} \cap \Pi(A) = \Pi(\bigcup(\mathcal{B}_0(A))$, and so $\Pi(A)/\Pi_{n-1} \cong \Pi(A)/\Pi(\bigcup \mathcal{B}_0(A))$ is $O(f)$ by Lemma 7.8. Since this applies to each component of Π_n, we see that Π_n/Π_{n-1} is $O(f)$.

Now setting $\Pi_\nu = \Pi$ and using the remark following Lemma 7.8, we see that $\Pi_\nu/\Pi_{\nu-1}$ is $O(f)$. Also, $\Pi_0 = \varnothing$, and so we have an increasing sequence of full subgraphs, $\varnothing = \Pi_0 \subseteq \Pi_1 \subseteq \cdots \subseteq \Pi_\nu = \Pi$, where Π_n/Π_{n-1} is $O(f)$ for all n. Applying Lemma 7.2, we see that Π is $O(f^\nu)$. But $g_\nu = f^\nu$ is a polynomial of degree ν. We have shown:

Lemma 7.9. *There is a sequence, $(g_\nu)_\nu$ of polynomials, g_ν of degree ν, such that any graph Π constructed in this way is $O(g_\nu)$.* □

Since $\theta : \Pi \to \Theta(M)$ is a uniform quasi-isometry, and since Π has uniformly bounded degree, it follows easily that the volume growth of $\Theta(M)$ about any point is bounded by some uniform polynomial, f_ν, of degree $\nu = \nu(\Sigma)$.

We have assumed that M is doubly degenerate, and pretended that Σ is a closed surface, but the general case proceeds in essentially the same way (see Section 8).

This proves Theorem 7.1.

8. The general case

In most of this paper, we have only dealt explicitly with the special case where Σ is a closed orientable surface, and M is orientable and without cusps. Moreover, we have mostly supposed that M is doubly degenerate. This has been mainly to simplify the exposition. The general case of a manifold admitting a type-preserving homotopy equivalence to a compact surface can be dealt with by fairly routine reinterpretations of various definitions and constructions as outlined below. In particular, Theorem 7.1 remains valid as stated in the general case.

Let M be a complete orientable hyperbolic 3-manifold admitting a homotopy equivalence to a compact surface Σ. We assume that this is *type-preserving*, that is, each boundary curve of Σ corresponds to a cusp of M. We write $X(\Sigma)$ for the set of homotopy classes of nonperipheral closed curves in Σ. We shall assume for the moment that Σ and M are orientable.

After fixing some Margulis constant, we have, as before, a set, $\mathcal{T}$, of Margulis tubes. In addition, we have a set, $\mathcal{P}$, of Margulis cusps. If $P \in \mathcal{P}$, then ∂P is a euclidean cylinder foliated by euclidean geodesic "longitudes" of fixed length. We write $N(M) = M \setminus \bigcup_{P\in\mathcal{P}} \operatorname{int} P$ for the *noncuspidal part* of M, and $\Theta(M) = N(M) \setminus \bigcup_{T\in\mathcal{T}} \operatorname{int} T$ for the *thick part* of M.

Let $\mathcal{P}_\partial(M)$ be the set of Margulis cusps that correspond to boundary components of Σ, and let $Q(M) = M \setminus \bigcup_{P\in\mathcal{P}_\partial(M)} \operatorname{int} P$. (Thus $\Theta(M) \subseteq N(M) \subseteq Q(M)$.) By tameness [Bonahon 1986], $Q(M)$ is homeomorphic to $\Sigma \times \mathbb{R}$. We refer to the ends $\Sigma \times [0, \infty)$ and $\Sigma \times (-\infty, 0]$ as the *positive* and *negative* ends of $Q(M)$. Note that

$\partial Q(M) = \bigcup_{P \in \mathcal{P}_\partial(M)} \partial P \equiv \partial\Sigma \times \mathbb{R}$. A *fibre* of $Q(M)$ is the image of a homotopy equivalence from Σ to $Q(M)$ where the preimage of $\partial Q(M)$ in Σ is precisely $\partial\Sigma$.

Let $\mathcal{P}_A(M) = \mathcal{P} \setminus \mathcal{P}_\partial(M)$. These are the *accidental* parabolic cusps of M. We can write $\mathcal{P}_A(M) = \mathcal{P}_+(M) \sqcup \mathcal{P}_-(M)$ depending on whether the cusp lies in the positive or negative end of $Q(M)$. Each $P \in \mathcal{P}_A(M)$ is homotopic to a curve $\alpha(P) \in X(\Sigma)$. The set $\{\alpha(P) \mid P \in \mathcal{P}_\pm(M)\}$ a multicurve in Σ, i.e. the elements are mutually disjoint. In particular, $\mathcal{P}_\pm(M)$ and hence $\mathcal{P}$ are finite.

A surface $\Phi \in \mathcal{F}$ is assumed to have the property that each boundary curve in Φ that is peripheral in Σ is equal to this boundary curve, and that all other boundary curves of Φ lie in $\operatorname{int}\Sigma$. As before, we can define an unknotted surface, $F \subseteq M$, where we assume that $F \cap \partial Q(M)$ are precisely the boundary curves of F that are peripheral in $Q(M)$. Again, we have $\phi(F) \in \mathcal{F} \setminus \{\Sigma\}$. We can similarly define a *thick surface*.

We need to modify the definitions of “horizontal surface” and “band”.

A *horizontal surface* is now an unknotted surface, $F \subseteq Q(M)$, such that there are two disjoint subsets, $\mathcal{T}_\partial(F)$ and $\mathcal{T}_I(F)$ of $\mathcal{T}$, satisfying (1)–(3) as before, and in addition, two disjoint subsets, $\mathcal{P}_\partial(F)$ and $\mathcal{P}_I(F)$ of $\mathcal{P}$ which satisfy (1′)–(3′), where $\mathcal{T}$, $\mathcal{T}_\partial(F)$ and $\mathcal{T}_I(F)$ are replaced by $\mathcal{P}$, $\mathcal{P}_\partial(F)$ and $\mathcal{P}_I(F)$. Condition (4) gets replaced by

(4′) $\partial F \subseteq \bigcup \mathcal{T}_\partial(F) \cup \bigcup \mathcal{P}_\partial(F)$.

Necessarily, $\mathcal{P}_I(F) \subseteq \mathcal{P}_A(M)$.

We similarly modify the definition of a band. It is now a thick surface, B, in $Q(M)$, with subsets $\mathcal{T}_\partial(B)$, $\mathcal{T}_I(B)$, $\mathcal{T}_+(B)$, $\mathcal{T}_-(B) \subseteq \mathcal{T}$ satisfying (1)–(4), (6) and (7), as before, together with subsets $\mathcal{P}_\partial(B)$, $\mathcal{P}_+(B)$, $\mathcal{P}_I(B) \in \mathcal{P}$ satisfying (1′), (2′), (4′), (6′) and (7′) where $\mathcal{T}$ gets replaced by $\mathcal{P}$ etc., and $\mathcal{P}_I(B) = \varnothing$. Condition (5) gets replaced by

(5′) $\partial_V B \subseteq \bigcup \mathcal{T}_\partial(B) \cup \bigcup \mathcal{P}_\partial(B)$.

As before, we assume that $\phi(B) \neq \Sigma$.

We necessarily have $\mathcal{P}_\pm(B) \subseteq \mathcal{P}_A(B)$ and $\mathcal{P}_-(B) \cap \mathcal{P}_+(M) \subseteq \mathcal{P}_+(B)$ and $\mathcal{P}_+(B) \cap \mathcal{P}_-(M) \subseteq \mathcal{P}_-(B)$. We say that B is *primitive* if $\mathcal{T}_+(B) \cap \mathcal{T}_-(B) = \mathcal{P}_+(B) \cap \mathcal{P}_-(B) = \varnothing$. In this case, $\mathcal{P}_+(B) \cap \mathcal{P}_-(M) = \mathcal{P}_-(B) \cap \mathcal{P}_+(M) = \varnothing$.

Let $Y = \operatorname{core}(M)$ be the convex core of M, and let ∂Y denote the boundary of Y in M. The inclusion of $\partial Y \cap \Theta(M)$ into ∂Y is a homotopy equivalence. Each component, F, of $\partial Y \cap N(M)$ is a horizontal surface with $\mathcal{P}_I(F) = \mathcal{T}_\partial(F) = \varnothing$. Moreover, F cuts $N(M)$ into two components, one of which, $C(F)$, homeomorphic to $F \times [0, \infty)$. We can refer to F, and hence the corresponding component of ∂Y, as *positive* or *negative* depending on whether $C(F)$ lies in the positive or negative end of $Q(M)$. We write $\partial_- Y$ (respectively $\partial_+ Y$) for the union of positive (negative) components, so that $\partial Y = \partial_+ Y \sqcup \partial_- Y$.

Now each Margulis tube, $T \in \mathcal{T}$, meets Y. Indeed, we can write

$$\mathcal{T} = \mathcal{T}_I(Y) \cup \mathcal{T}_-(Y) \cup \mathcal{T}_+(Y)$$

with $\mathcal{T}_I(Y) \cap (\mathcal{T}_I(Y) \cup \mathcal{T}_+(Y)) = \varnothing$, so that for all $T \in \mathcal{T}_I(Y)$, $T \subseteq Y$, and for all $T \in \mathcal{T}_\pm(T)$, $\partial_\pm Y$ meets T in an annulus.

Let's first consider the case where M is geometrically finite. This means that $Y \cap \Theta(M)$ is compact, so that $\mathcal{T}$ is finite, and $Y \cap N(M)$ is compact. Indeed we can find disjoint horizontal fibres S_+ and S_- of $Q(M)$ such that

$$S_\pm \cap N(M) = \partial_\pm Y \cap N(M).$$

Now S_+ and S_- bound a compact region, K, in $Q(M)$. In fact, K, is like a band in $Q(M)$, with $\mathcal{T}_I(K) = \mathcal{T}_I(Y)$, $\mathcal{T}_\pm(K) = \mathcal{T}_\pm(Y)$, $\mathcal{T}_\partial(K) = \varnothing$, $\mathcal{P}_\pm(K) = \mathcal{P}_\pm(Y)$ and $\mathcal{P}_\partial(K) = \mathcal{P}_\partial(M)$, except that $\phi(K) = \Sigma$, which we have disallowed.

The statement of Theorem 0 is similar to that given in Section 2. We construct a nested set, $\mathcal{B}$, of bands satisfying (B1)–(B6) of Section 5. This time, we assume that each band lies in the interior of Y. We let $\mathcal{A} \subseteq \mathcal{B}$ be the set of outermost bands. These bands satisfy (A1), (A2) and (A4)–(A9) of Section 2. Property (A3) should now say that $L(\partial T \cap Y, \mathcal{A}) \leq L_0$ for all $T \in \mathcal{T}$, and $L(\partial P \cap Y, \mathcal{A}) \leq L_0$ for all $P \in \mathcal{P}_A(M)$. To the statement of (A3′), we should add that $L(\partial P \cap B, \mathcal{B}(B)) \leq L_0$ for all $P \in \mathcal{P}_A(M)$.

The case where there are no accidental parabolics — $\mathcal{P}_A(M) = \varnothing$ — is similar. In this case, each of $\partial_+ Y$ and $\partial_- Y$ is either empty or a horizontal fibre, and so we have a division into geometrically finite, singly degenerate and doubly degenerate cases. The statement of Theorem 0 is as for the geometrically finite case above.

For a manifold with accidental parabolics that is not geometrically finite, the situation a bit more complicated. One way of dealing with it is to allow for "long bands" where one of the horizontal boundary components may be at infinity.

More precisely, a semi-infinite thick surface, B, is the image of a proper embedding of $\Phi \times [0, \infty)$ into $Q(M)$, where $\Phi \in \mathcal{F}$. We write $\partial_H B$ for the image of $\Phi \times \{0\}$ and $\partial_V B$ for the image of $\partial\Phi \times [0, \infty)$. A *long band* is now a semi-infinite thick surface B, with $\partial_V B \subseteq \bigcup \mathcal{P}$ and with $\partial_H B$ a horizontal surface.

We now allow $\mathcal{B}$ to contain (a necessarily finite number of) long bands. We can assume that $\mathcal{B}$ satisfies (B1)–(B6). For a long band, $\mathcal{B}$, (B4) is redundant and (B5) means that if F is parallel to B, then $H(\langle F, \partial_H B\rangle) \leq H_0 + 2D_0$. If $\mathcal{A}$ is the set of outermost bands, then conditions (A1)–(A9) are satisfied, with (A3) and (A3′) modified as above. Indeed, if $P \in \mathcal{P}_A(M)$, then $\partial P \cap Y \setminus \bigcup \mathcal{A}$ is compact.

Let $\mathcal{C} \subseteq \mathcal{B}$ be the set of innermost long bands. These are disjoint. If $C \in \mathcal{C}$, and $P \in \mathcal{P}$, then $P \cap C \subseteq \partial P$, otherwise we could find smaller long bands contained in $\mathcal{C}$. Thus $C \subseteq N(M)$. Let F_+ be the union of $\partial_H C$ as C varies over positive bands in $\mathcal{C}$. We can find a horizontal fibre, S_+, of $Q(M)$ such that $S_+ \cap N(M) =$

$(F_+ \cup \partial_+ Y) \cap N(M)$. We can similarly find a disjoint fibre, S_-. Let K be the compact region of S_+ and S_-.

We see that K behaves like the compact region K constructed in the geometrically finite case. (Note $K \cap N(M)$ need not be connected.) Similarly, each band of $\mathscr{C}$ behaves like the convex core of a singly degenerate manifold with smaller base surface. Thus, in some sense, the general case is a union of geometrically finite and singly degenerate cases.

Finally, we remark that the nonorientable case can also be similarly accounted for. In this case, Margulis tubes may be solid tori, and boundaries of Margulis cusps may be Möbius bands. Also, there may be no canonical choice of "positive" or "negative" boundaries of bands.

References

[Bell and Fujiwara 2005] G. Bell and K. Fujiwara, "The asymptotic dimension of the curve graph is finite", preprint, 2005. math.GT/0509216

[Bonahon 1986] F. Bonahon, "Bouts des variétés hyperboliques de dimension 3", *Ann. of Math.* (2) **124**:1 (1986), 71–158.

[Bonahon and Otal 1988] F. Bonahon and J.-P. Otal, "Variétés hyperboliques à géodésiques arbitrairement courtes", *Bull. London Math. Soc.* **20** (1988), 255–261.

[Bowditch 2003] B. H. Bowditch, "Tight geodesics in the curve complex", preprint, University of Southampton, 2003. To appear in *Inv. Math.*

[Bowditch 2005a] B. H. Bowditch, "End invariants of hyperbolic 3-manifolds", preprint, University of Southampton, 2005.

[Bowditch 2005b] B. H. Bowditch, "Geometric models for hyperbolic 3-manifolds", preprint, University of Southampton, 2005.

[Brock 2001] J. F. Brock, "Iteration of mapping classes and limits of hyperbolic 3-manifolds", *Invent. Math.* **143**:3 (2001), 523–570.

[Brock et al. 2004] J. Brock, R. D. Canary, and Y. Minsky, "Classification of Kleinian surface groups, II: The ending lamination conjecture", preprint, 2004. math.GT/0412006

[Brock et al. $\geq$ 2007] J. Brock, R. D. Canary, and Y. Minsky, "The classification of finitely-generated Kleinian groups". in preparation.

[Brown 1966] E. M. Brown, "Unknotting in $M^2 \times I$", *Trans. Amer. Math. Soc.* **123** (1966), 480–505.

[Canary 1996] R. D. Canary, "A covering theorem for hyperbolic 3-manifolds and its applications", *Topology* **35**:3 (1996), 751–778.

[Freedman et al. 1983] M. Freedman, J. Hass, and P. Scott, "Least area incompressible surfaces in 3-manifolds", *Invent. Math.* **71**:3 (1983), 609–642.

[Gromov 2007] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Prog. Math. **152**, Birkhäuser, Boston, 2007.

[Kida 2005] Y. Kida, "The mapping class group from the viewpoint of measure equivalence theory", preprint, 2005. To appear in *Mem. Amer. Math. Soc.* math/0512230

[Minsky 1999] Y. N. Minsky, "The classification of punctured-torus groups", *Ann. of Math.* (2) **149**:2 (1999), 559–626.

[Minsky 2002] Y. N. Minsky, "The classification of Kleinian surface groups, I: Models and bounds", preprint, SUNY Stony Brook, 2002.

[Otal 1995] J.-P. Otal, "Sur le nouage des géodésiques dans les variétés hyperboliques", *C. R. Acad. Sci. Paris Sér. I Math.* **320**:7 (1995), 847–852.

[Otal 2003] J.-P. Otal, "Les géodésiques fermées d'une variété hyperbolique en tant que nœuds", pp. 95–104 in *Kleinian groups and hyperbolic 3-manifolds* (Warwick, 2001), London Math. Soc. Lecture Note Ser. **299**, Cambridge Univ. Press, Cambridge, 2003.

[Soma 2003] T. Soma, "Geometric limits of quasifuchsian groups", preprint, Tokyo Denki Univ., 2003.

[Thurston 1979] W. P. Thurston, "The geometry and topology of three-manifolds", lecture notes, Princeton University, 1979, Available at http://msri.org/publications/books/gt3m.

Received January 16, 2006.

BRIAN H. BOWDITCH
MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
GREAT BRITAIN
http://www.warwick.ac.uk/~masgak

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

CLASSIFICATION OF FIBER SURFACES OF GENUS 2 WITH AUTOMORPHISMS ACTING TRIVIALLY IN COHOMOLOGY

JIN-XING CAI

Let S be a complex nonsingular projective surface of general type with a fibration of genus 2, and let $G \subset \operatorname{Aut} S$ be a nontrivial subgroup of automorphisms of S, inducing trivial actions on $H^2(S, \mathbb{Q})$. We give a classification for pairs (S, G) from the point of view of moduli. Consequently, we show that there exist surfaces S of general type (with p_g arbitrary large) with an involution acting trivially on $H^i(S, \mathbb{Z})$ for all i.

1. Introduction

Let S be a complex minimal nonsingular projective surface of general type, and let $G \subset \operatorname{Aut} S$ be a *nontrivial* subgroup of automorphisms of S inducing trivial actions on $H^2(S, \mathbb{Q})$. Peters [1979] proved that, if the canonical linear system $|K_S|$ is basepoint free, then either $K_S^2 = 8\chi(\mathcal{O}_S)$ or $K_S^2 = 9\chi(\mathcal{O}_S)$. Recently, we showed that $|G| \leq 4$ if $\chi(\mathcal{O}_S) > 188$ [Cai 2004]. When S has a fibration of genus 2, we have a numerical classification for pairs (S, G):

Theorem 1.1 [Cai 2006a; 2006b]. *Let S, G be as above. Assume that S has a relatively minimal fibration of genus 2 and $\chi(\mathcal{O}_S) \geq 5$. Then $|G| = 2$, and either*

(i) $K_S^2 = 4\chi(\mathcal{O}_S) - 4a$ $(a = 0,\ 1)$, *or*

(ii) $K_S^2 = 8\chi(\mathcal{O}_S) - 6b$ $(b = 0,\ 1,\ 2)$.

There are some examples in [Cai 2006a; 2006b] to show that such pairs (S, G) exist, besides the well known ones (products of two hyperelliptic curves). An interesting question is whether it could be possible to classify all possible pairs (S, G) in Theorem 1.1.

In this note we give a classification for pairs (S, G) in Theorem 1.1 from the more general point of view of moduli. Roughly speaking, our main result is this (see Theorems 2.5 and 4.7 for precise statements):

Theorem 1.2. *Let S, G be as in Theorem 1.1.*

MSC2000: primary 14J50; secondary 14J29.
Keywords: surfaces of general type, automorphism groups, fibrations.

(i) *If S is as in Theorem 1.1(i), then S is birationally equivalent to a double cover of certain elliptic fiber bundle. The configuration of the ramification divisor of this covering is determined.*

(ii) *If S is as in Theorem 1.1(ii) with $b=0$, then $S \simeq (F \times \tilde{C})/\tilde{G}$, where F and $\tilde{C}$ are curves of genus $g(F)=2$, $g(\tilde{C}) \geq 2$, and $\tilde{G}$ is one of the following groups*: $\mathbb{Z}/m\mathbb{Z}$ ($m \leq 10$, $m \neq 7, 9$), $(\mathbb{Z}/2\mathbb{Z})^2$, $\mathbb{Z}/6\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$, D_8 (*the dihedral group of order* 8); *a complete description for the action of $\tilde{G}$ on $F \times \tilde{C}$ is given.*

We note that, for K3 and Enriques surfaces S, $\operatorname{Aut} S$ acts faithfully on $H^2(S, \mathbb{Z})$ (see [Burns and Rapoport 1975; Ueno 1976]). As an interesting consequence of Theorem 2.5, we show that the analogous question for surfaces of general type has a negative answer:

Theorem 1.3 (Corollary 2.11). *Let $n \geq 3$ be an integer. There exist an infinite series of surfaces S_n of general type with $K_{S_n}^2 = 4n$, $p_g(S_n) = n$, $q(S_n) = 1$ admitting an involution acting trivially on $H^i(S_n, \mathbb{Z})$ for all i.*

We work over the complex number field and use standard notation as exemplified by [Barth et al. 1984]. We also use freely the notation from [Cai 2006a; 2006b].

2. Surfaces whose canonical map being composite with a pencil

2.1. Let S be a complex nonsingular projective surface of general type with $p_g(S)$ at least 3 and let $f : S \to C$ be a relatively minimal fibration of genus 2. Consider a nontrivial subgroup $G \subset \operatorname{Aut} S$ of automorphisms of S inducing trivial actions on $H^2(S, \mathbb{Q})$. In this section, we assume that the canonical map Φ_S of S is composite with a pencil. By [Cai 2006a, Theorem 3.2], we have $|G| = 2$, the generator σ of G is a bielliptic involution of f (that is, $f \circ \sigma = f$, and for a general fiber F of f, $\sigma_{|F}$ is a bielliptic involution of F), and S has numerical invariants

(2.1.1) $K_S^2 = 4\chi(\mathcal{O}_S)$ and $q(S) = g(C) = 1$, or

(2.1.2) $K_S^2 = 4\chi(\mathcal{O}_S) - 4$, $q(S) = 1$ and $g(C) = 0$.

The hyperelliptic involutions of smooth fibers of f glue together to give a birational C-involution τ of S, which is everywhere defined by the uniqueness of the minimal model of f. We call τ the hyperelliptic involution of $f : S \to C$. Let $\lambda = \sigma \circ \tau$. Clearly λ is a bielliptic involution of f. We have a commutative diagram

$$\begin{tikzcd}
\tilde{S} \arrow[r, "\tilde{\pi}"] \arrow[d, "\rho"'] & \tilde{S}/\tilde{\lambda} \arrow[r, "\alpha"] \arrow[d] & T \arrow[dl, "p"] \\
S \arrow[r, "f"] & C &
\end{tikzcd}$$

where ρ is the blowup of all isolated fixed points of λ, $\tilde{\lambda}$ is the induced involution on $\tilde{S}$, α is the blowdown of all -1-curves contained in fibers of $\tilde{S}/\tilde{\lambda} \to C$, and p is the induced relatively minimal elliptic fibration.

We can describe $p : T \to C$ explicitly:

Proposition 2.2. *Let E_2 be an elliptic curve, and set $E_4 = \mathbb{C}/(\mathbb{Z} + i\mathbb{Z})$ and $E_3 = E_6 = \mathbb{C}/(\mathbb{Z} + \xi\mathbb{Z})$, for ξ a primitive third root of unit.*

(i) *If S is as in* (2.1.1), *then C is an elliptic curve, and*

$$(p : T \to C) \simeq (T_d := (C' \times E_d)/\mathbb{Z}_d \to C'/\mathbb{Z}_d)$$

for some $d \in \{2, 3, 4, 6\}$, where C' is an elliptic curve and $\mathbb{Z}_d$ acts on $C' \times E_d$ via a product action: $\mathbb{Z}_d$ acts on C' as a translation of order d such that $C'/\mathbb{Z}_d \simeq C$, and $\mathbb{Z}_d$ acts on E_d by (1) *$e \mapsto -e$ if $d = 2$;* (2) *$e \mapsto \xi e$ if $d = 3$;* (3) *$e \mapsto ie$ if $d = 4$;* (4) *$e \mapsto -\xi e$ if $d = 6$.*

Moreover, $K_T = p^*\eta$, where $\eta \in \operatorname{Pic}^0 C$, which determines the étale cover $C' \to C$.

(ii) *If S is as in* (2.1.2), *then $C = \mathbb{P}^1$, $T = C \times E$ and p is the projection to the first factor, where E is an elliptic curve.*

Proof. By [Cai 2006a, Proposition 4.12] and its proof, $p\colon T \to C$ is an elliptic fiber bundle with a section. By the proof of Theorem 3.2 of the same reference, we have $q(T) = g(C) = 1$ if S is as in (2.1.1), and $q(T) = 1$, $g(C) = 0$ if S is as in (2.1.2). Note that $p_g(T) = 0$. Now the proposition follows from the well-known result of Bagnera and de Franchis on the classification of bielliptic surfaces (see [Beauville 1983, VI, 20], for example). □

Proposition 2.3. *Let T_d be as in Proposition 2.2. Then $H_1(T_d, \mathbb{Z})_{\mathrm{tor}} \simeq \mathbb{Z}_2 \times \mathbb{Z}_2$, $\mathbb{Z}_3$, $\mathbb{Z}_2$, 0 if $d = 2, 3, 4, 6$, respectively.*

See [Iitaka 1971; Suwa 1969; Serrano 1991] for proofs.

Notation 2.4. Let $p : T \to C$ be a fiber surface and $\Delta \subset T$ a bisection of T, that is, an irreducible curve with $\Delta P = 2$, where P is a fiber of p. We say that a point $t \in \Delta$ is a ramification point of $p_{|\Delta} : \Delta \to C$ if t is in the image of the set of ramification points of $p_{|\Delta} \circ \phi : \tilde{\Delta} \to \Delta \to C$ under ϕ, where $\phi : \tilde{\Delta} \to \Delta$ is the normalization of Δ.

For any point $t \in \Delta$, let $l(t; \Delta)$ be the number of times we must blow up $t \in T$ and its infinitely near points to get the strict transform of Δ being nonsingular at the inverse image of t.

For any two curves D, D' and $t \in D \cap D'$, we denote by $I(D, D'; t)$ the intersection number of D and D' at the point t.

Theorem 2.5. *Let $f\colon S \to C$, $p : T \to C$, $\tilde{\pi}$, and α be as in 2.1. Let $\pi : S' \to T$ be the Stein factorization of $\alpha \circ \tilde{\pi}$, and let (B, θ) be the singular double cover data corresponding to π. Then (B, θ) has the following properties*:

(i) $\theta = C_1 + p^*D$, *where C_1 is a section of p and D is a divisor on C of degree $n := p_g(S) \geq 3$,*

(ii) $B = \Delta + \sum_{i=1}^m p^*c_i$, *where $\Delta \in |2C_1 + p^*(2D - \sum_{i=1}^m c_i)|$ is a bisection of p and c_i $(i = 1, \dots, m)$ are different points of C,*

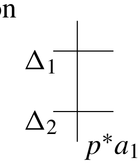
(iii) $\Delta \cap C_1$ *is contained in the set of ramification points of $p_{|\Delta}$. As a set, $\Delta \cap C_1 = \{t_1, \cdots, t_m\}$, where $t_i = p^*c_i \cap C_1$. For any i, $I(\Delta, C_1; t_i) = 2l(t_i; \Delta) + 1$. So $\sum_{i=1}^m l(t_i; \Delta) = n - m$.*

Conversely, let $p : T \to C$ be as in Proposition 2.2, and let (B, θ) be the singular double covers data satisfying conditions (*i*)–(*iii*) *above. Let $\pi\colon S' \to T$ be the double cover corresponding to (B, θ). Let S'' be the desingularization of S', and $f'\colon S'' \to C$ the induced fibration. Let $f\colon S \to C$ be the relatively minimal fibration of f'. Denote by τ the hyperelliptic involution of f, and λ the involution corresponding to the double cover π. Let $\sigma = \tau \circ \lambda$. Then S is as in* (2.1.1) (*resp.* (2.1.2)) *with $p_g(S) = n$ if T is as in* (*i*) (*resp.* (*ii*)) *of Proposition 2.2 and σ acts trivially on $H^2(S, \mathbb{Q})$.*

Proof. We assume that T is as in Proposition 2.2(i). The proof of the other case is similar and is left to the reader. Since B has no essential singularities, by the formula for double covers, we have $h^0(K_T \otimes \theta) = n$. Note that $p : T \to C$ is a fiber bundle, and $(K_T \otimes \theta)P = 1$ for a fiber P of p. We have $K_T \otimes \theta \equiv C_1 + p^*D'$, where C_1 is a section of p and D' is an effective divisor on C. Clearly C_1 is the fixed part of $|C_1 + p^*D'|$. So $\deg D' = h^0(D') = h^0(C_1 + p^*D') = n$. Note that $K_T = p^*\eta$, where η is as in Proposition 2.2. So $\theta = C_1 + p^*D$, where $D = D' \otimes \eta$ is a divisor on C of degree n.

Since B is a reduced divisor, we may write $B = \Delta + \sum_{i=1}^m p^*c_i$, where Δ is a reduced horizontal divisor with respect to p, $m \geq 0$, and c_i $(i = 1, \dots, m)$ are different points of C.

2.6. We show that Δ is irreducible. Otherwise, $\Delta = \Delta_1 + \Delta_2$, where Δ_j are sections of p. Clearly $\Delta_1\Delta_2 = 0$. So $m > 0$. Then locally around p^*c_1 the branch locus B of π has the configuration



So $(p \circ \pi)^*c_1$ is a multiple fiber and S' has two rational double points on it, and hence f^*c_1 is a fiber of type (b_0). This contradicts [Cai 2006a, Lemma 4.7(ii)].

Lemma 2.7. *If $t \in \Delta \cap C_1$, then t is a ramification point of $p_{|\Delta}$, and*

$$I(\Delta, C_1; t)) = 2l(t; \Delta) + 1,$$

where $l(t; \Delta)$ is as in Notation 2.4.

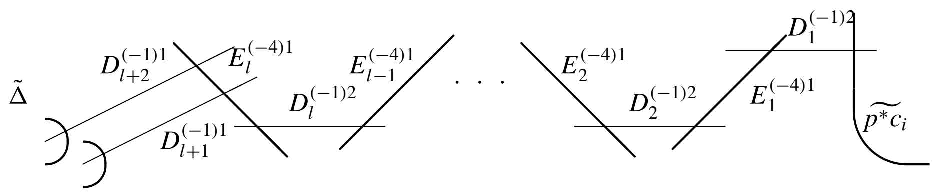
Proof. let $c = p(t)$ and $l = l(t; \Delta)$. First we assume that t is a smooth point of Δ. If t is not a ramification point of $p_{|\Delta}$, then $p^*c \cap \Delta$ consists of two different points, t and t'. We have $t + t' - 2t \equiv \Delta_{|p^*c} - 2C_{1|p^*c} \equiv p^*(2D - \sum_{i=1}^m c_i)_{|p^*c} \equiv 0$. This implies $t \equiv t'$ on p^*c, which is a contradiction since p^*c is not rational.

Now we may assume that t is a singular point of Δ. If $c \neq c_i$ for any i, then $\mathrm{mult}_t B = 2$. Let $\hat{\rho} : \hat{T} \to T$ be the blowing up at t, and E the exceptional curve. For any irreducible curve Z in T, we denote $\hat{Z}$ the strict transform of Z in $\hat{T}$. Set

$$\hat{B} = \hat{\rho}^* B - 2E, \quad \hat{\theta} = \hat{\rho}^*\theta - E = \hat{C}_1 + \hat{\rho}^* p^* D.$$

Let $\hat{\pi} : \hat{S} \to \hat{T}$ be the double cover corresponding to $(\hat{B}, \hat{\theta})$. Clearly $\alpha \circ \tilde{\pi}$ (notation as in 2.1) factors through $\hat{\pi}$. Since C_1 and p^*c meet transversally only in one point t, we have $\widehat{C_1} \cap \widehat{p^*c} = \varnothing$. This implies $\hat{\theta}_{|\widehat{p^*c}}$ is trivial. So $\hat{\pi}^*\widehat{p^*c}$ has two disconnected components, and hence f^*c is of type (a_k). This contradicts [Cai 2006a, Lemma 4.6].

So we can assume $c = c_i$ for some i. If $t \in \Delta$ is not a ramification point of $p_{|\Delta}$, then $(p \circ \alpha)^* c$ has the following configuration:



where $\tilde{\Delta}$ and $\widetilde{p^*c_i}$ are the strict transforms of Δ and p^*c_i, thick lines mean branch locus of $\tilde{\pi}$, and superscript numbers without brackets are multiplicities and superscript numbers within brackets denote self-intersections. This implies f^*c_i is of type (b_{2l}), which is a contradiction by [Cai 2006a, Lemma 4.7(ii)].

Now $t \in \Delta$ is a ramification point of $p_{|\Delta}$. Let $H = (\alpha \circ \tilde{\pi})^* C_1$. By [Cai 2006a, 4.8, 4.11 and 4.12], we have $(f \circ \rho)_{|H} : H \to C$ is étale. So the strict transform $\tilde{C}_1$ of C_1 in $\tilde{S}/\tilde{\lambda}$ does not meet the branch locus of $\tilde{\pi}$. This implies $I(\Delta, C_1; t) = 2l + 1$ by a standard calculation; see, for instance, [Hartshorne 1977, Chapter V, Propositions 3.2 and 3.6]. □

By the proof of Lemma 2.7, the image of $\Delta \cap C_1$ under p is contained in the set $\{c_1, \ldots, c_m\}$. Now suppose there is a point $c_i \in \{c_1, \ldots, c_m\} \setminus p(\Delta \cap C_1)$. If $p^*c_i \cap \Delta$ consists of two points, then $p_{|\Delta}$ is étale at c_i and we get a contradiction as in 2.6. Hence $p^*c_i \cap \Delta$ is a single point. By the choice of c_i, $p^*c_i \cap \Delta \notin C_1$.

So $p^*c_i \cap C_1 \neq p^*c_i \cap \Delta$, and hence $p^*c_i \cap C_1$ must be a smooth point of B. This implies the strict transform $\tilde{C}_1$ of C_1 does meet the branch locus of $\tilde{\pi}$. This is impossible since $H \to C$ is étale.

Now we prove the converse of the theorem. Let T be as in (i) of Proposition 2.2, and let (B, θ) be the singular double cover data satisfying conditions (i)–(iii) in Theorem 2.5. Let $\pi : S' \to T$ be the double cover corresponding to (B, θ). Then S' has only canonical singularities. Let $\epsilon : S \to S'$ be the minimal desingularization. We have

$$K_S = (\pi \circ \epsilon)^*(p^*(\eta + D) + C_1).$$

So S has the following numerical invariants:

$$K_S^2 = 4n,\ p_g(S) = n,\ q(S) = 1.$$

Now $f := p \circ \pi \circ \epsilon : S \to C$ is a fibration of genus 2. Denote by τ the hyperelliptic involution of f, and by λ the involution of S corresponding to π. Then $\langle \lambda, \tau \rangle \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$. Take $\sigma = \tau \circ \lambda$. Now the result follows by the following lemma. □

Lemma 2.8. *The involution σ acts trivially on $H^i(S, \mathbb{Q})$ for all i.*

The idea of the proof of Lemma 2.8 is to analyze the action of σ around the singular fibers of f, and to apply the topological Lefschetz formula to σ. The proof is longer and is postponed until the next section; see also [Cai 2006a, 3.3] for a proof in the special case when the bisection $\Delta < B$ is smooth.

Remark 2.9. Let Δ be as in Theorem 2.5. If Δ is smooth, then $l(t_i; \Delta) = 0$ for all i and hence $m = n$. In this case, by the proof of Lemma 2.7, the points in $\Delta \cap C_1$ are necessarily ramification points of $p_{|\Delta}$. So the only condition for (S, σ) being as in 2.1 is that the n fibers p^*c_i contained in B pass through the n points of $\Delta \cap C_1$.

Corollary 2.10. (i) *The moduli space $\mathcal{M}$ of surfaces (S, σ) as in* (2.1.1) *with $p_g(S) = n$ has four irreducible connected components. Among them one has dimension $2n + 1$ and the others have dimensions $2n$.*

(ii) *The moduli space $\mathcal{M}'$ of surfaces (S, σ) as in* (2.1.2) *with $p_g(S) = n$ is irreducible and of dimension $2n - 1$.*

Proof. We prove (i); the proof of (ii) is similar and is left to the reader. By Theorem 2.5, $\mathcal{M}$ is a disjoint union $\mathcal{M}_2 \cup \mathcal{M}_3 \cup \mathcal{M}_4 \cup \mathcal{M}_6$, where $\mathcal{M}_d = \{S \in \mathcal{M} \mid T \simeq T_d\}$, for T_d is as in Proposition 2.2. Let $\mathcal{B}_z \in |2\theta|$ be a flat family of curves such that $\mathcal{B}_0$ is the branch locus B of $\pi : S' \to T$ and $\mathcal{B}_1$ is smooth. Let $\mathcal{S}_z$ be the flat family of surfaces corresponding to the double cover data $(\mathcal{B}_z, \theta)$. Since the branch locus $\mathcal{B}_1$ of $\mathcal{S}_1 \to T$ is ample, we have $\pi_1(\mathcal{S}_1) \simeq \pi_1(T)$ by [Cornalba 1981]. Since $\mathcal{B}_0 = B$ has no essential singularities, $S' = \mathcal{S}_0$ has only rational double points. By [Atiyah 1958], the minimal desingularization S of $\mathcal{S}_0$ is diffeomorphic to $\mathcal{S}_1$. Hence we have $\pi_1(S) \simeq \pi_1(T)$. By Proposition 2.3, the sets $\mathcal{M}_d$ are open. Given

T_d, for generic $[S] \in \mathcal{M}_d$, S is determined by (B, θ), where $\theta = C_1 + p^*D$, D is a divisor of degree n on C, $B = \Delta + \sum_{i=1}^{n} p^*c_i$, $\Delta \in |2C_1 + p^*D|$ (cf. Remark 2.9). Up to automorphisms of T_d, C_1 is uniquely determined. Given a smooth curve $\Delta \in |2C_1 + p^*D|$, the choice of θ is unique up to a torsion element of order 2 of $\mathrm{Pic}^0\, T_d$. Clearly Δ depends on $h^0(2C_1 + p^*D) - 1 = 2n - 1$ (by Riemann–Roch) parameters. Note that T_d depends on two parameters if $d = 2$, and on one if $d = 3,\ 4,\ 6$. So the dimension of $\mathcal{M}_d$ is $2n + 1$ if $d = 2$, and $2n$ if $d = 3,\ 4,\ 6$. □

Corollary 2.11. *Let (S, σ) be as in* (2.1.1). *If $S \in \mathcal{M}_6$, where $\mathcal{M}_6$ is as in the proof of Corollary 2.10, then the involution σ acts trivially on $H^i(S, \mathbb{Z})$ for all i.*

Proof. If $S \in \mathcal{M}_6$ we have $\pi_1(S) \simeq \pi_1(T)$ by the proof of Corollary 2.10, and hence $H_1(S, \mathbb{Z})_{\mathrm{tor}} = 0$ by Proposition 2.3. By the Poincaré duality for the torsion part of homology, we have $H^2(S, \mathbb{Z})_{\mathrm{tor}} = 0$. Hence $H^*(S, \mathbb{Z})$ is torsion-free, and the result follows from Lemma 2.8. □

3. Proof of Lemma 2.8

We keep the notation of Theorem 2.5. Since $q(S) = g(C)$, by Hodge theory, σ acts trivially on $H^1(S, \mathbb{Q})$. To check that the involution σ acts trivially on $H^2(S, \mathbb{Q})$, we analyze the action of σ around the singular fibers of f. Let t_j $(j = 1, \ldots, u)$ be the ramification points of $p_{|\Delta}$. After suitable reindexing, we may assume that $\{t_1, \ldots, t_m\} = \Delta \cap C_1$ as a set. Let t_{u+k} $(k = 1, \ldots, v)$ be the singular points of $\Delta \setminus \{t_j | 1 \le j \le u\}$. Set $l_j = l(t_j; \Delta)$, where $l(t_j; \Delta)$ is as in Notation 2.4. We have $l_j \ge 0$ for $j = 1, \ldots, u$, and $l_j \ge 1$ for $j = u+1, \ldots, u+v$. By the definition of l_j, we have

$$p_a(\Delta) = g(\tilde{\Delta}) + \sum_{j=1}^{u+v} l_j,$$

where $\phi : \tilde{\Delta} \to \Delta$ is the normalization of Δ. Applying the Hurwitz formula to $p_{|\Delta} \circ \phi$, we get

$$2g(\tilde{\Delta}) - 2 = u.$$

By the adjunction formula,

$$2p_a(\Delta) - 2 = \Big(2C_1 + p^*\Big(2D - \sum_{j=1}^{m} c_j\Big)\Big)^2 = 4(2n - m).$$

Combining these three equalities, we have

$$4m + u + 2\sum_{j=1}^{u+v} l_j = 8n. \tag{3.0.1}$$

Let $\varrho \colon T' \to T$ be the morphism composed of l_j times blow-ups of t_j and its infinitely near points $(j = 1, \ldots, u+v)$. The exceptional divisor $\varrho^*(t_j)$ equals

$\sum_{l=1}^{l_j} E'_{jl}$, where E'_{jl} is the exceptional curve corresponding to the $(l-1)$-th near points of t_j. Then the strict transform Δ' of Δ is smooth, and for $j = 1, \dots, u$, Δ' meets E'_{jl_j} in one point t'_j and is tangent to it there. Let $\varrho' \colon T'' \to T'$ be the blow-up of t'_j ($j = 1, \dots, u$) and $s_{jl} := E'_{j,l-1} \cap E'_{jl}$ ($j = 1, \dots, m$, $l = 1, \dots, l_j$) (for convenience, here we set $E'_{j0} = (p^*c_j)'$). Let $E''_{j,l_j+1} = \varrho'^*(t'_j)$ and $D''_{jl} = \varrho'^*(s_{jl})$ be the exceptional curves. Then E''_{j,l_j+1} and the strict transform Δ'' of Δ' meet transversely at point t''_j. Let

$$\mu \colon \tilde{T} \to T''$$

be the blow-up of t''_j ($j = 1, \dots, u$). Let $\tilde{E}_{j,l_j+2} = \mu^*(t''_j)$ ($j = 1, \dots, u$) be the exceptional curves. For any irreducible curve Y in T, we denote by Y', Y'' and $\tilde{Y}$ the strict transform of Y in T', T'' and $\tilde{T}$, respectively. Set

$$\begin{aligned} \tilde{B} &:= \mu^*\Big(\varrho'^*\Big(\varrho^* B - 2\sum_{j=1}^{u+v}\sum_{l=1}^{l_j} l E'_{jl}\Big) - 2\sum_{j=1}^{m} E''_{j,l_j+1} - 2\sum_{j=1}^{m}\sum_{l=1}^{l_j} D''_{jl}\Big) - 2\sum_{j=1}^{u} \tilde{E}_{j,l_j+2} \\ &= \tilde{\Delta} + \sum_{j=1}^{m} \widetilde{p^*c}_j + \sum_{j=1}^{m}\sum_{l=1}^{l_j} \tilde{E}'_{jl} + \sum_{j=m+1}^{u} \tilde{E}''_{j,l_j+1} \end{aligned}$$

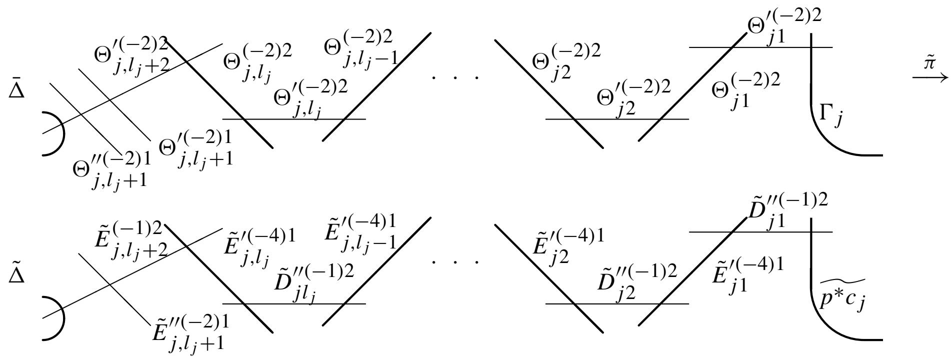
$$\begin{aligned} \tilde{\theta} &:= \mu^*\Big(\varrho'^*\Big(\varrho^* \theta - \sum_{j=1}^{u+v}\sum_{l=1}^{l_j} l E'_{jl}\Big) - \sum_{j=1}^{m} E''_{j,l_j+1} - \sum_{j=1}^{m}\sum_{l=1}^{l_j} D''_{jl}\Big) - \sum_{j=1}^{u} \tilde{E}_{j,l_j+2} \\ &= (\varrho \circ \varrho' \circ \mu)^* D + \tilde{C}_1 - \sum_{j=m+1}^{u+v}\sum_{l=1}^{l_j} l \tilde{E}'_{jl} - \sum_{j=1}^{m}\sum_{l=1}^{l_j} \widetilde{D}''_{jl} \\ &\qquad - \sum_{j=m+1}^{u} l_j \tilde{E}''_{j,l_j+1} - \sum_{j=1}^{m} \tilde{E}_{j,l_j+2} - \sum_{j=m+1}^{u} (2l_j+1)\tilde{E}_{j,l_j+2}. \end{aligned}$$

We have $\tilde{B}$ is a smooth divisor on $\tilde{T}$, and $\tilde{B} \equiv 2\tilde{\delta}$. Let $\tilde{\pi} \colon \tilde{S} \to \tilde{T}$ be the morphism associated with the double cover data $(\tilde{B}, \tilde{\delta})$. By the canonical resolution [Persson 1978], we have a commutative diagram

$$\begin{tikzcd} & \tilde{S} \arrow[r, "\tilde{\pi}"] \arrow[dl, "\varepsilon"'] \arrow[d, "\beta"] & \tilde{T} \arrow[d, "\gamma := \varrho \circ \varrho' \circ \mu"] \\ S \arrow[r, "\epsilon"] & S' \arrow[r, "\pi"] & T, \end{tikzcd}$$

where β is a desingularization of S', and ε is the contraction of -1-curves on $\tilde{S}$.

Clearly f has only $u+v$ singular fibers f^*c_j ($j = 1, \dots, u+v$). For $j = 1, \dots, m$, locally around a singular fiber, $\tilde{\pi} \colon (f \circ \varepsilon)^* c_j \to (p \circ \gamma)^* c_j$ has the following configurations:

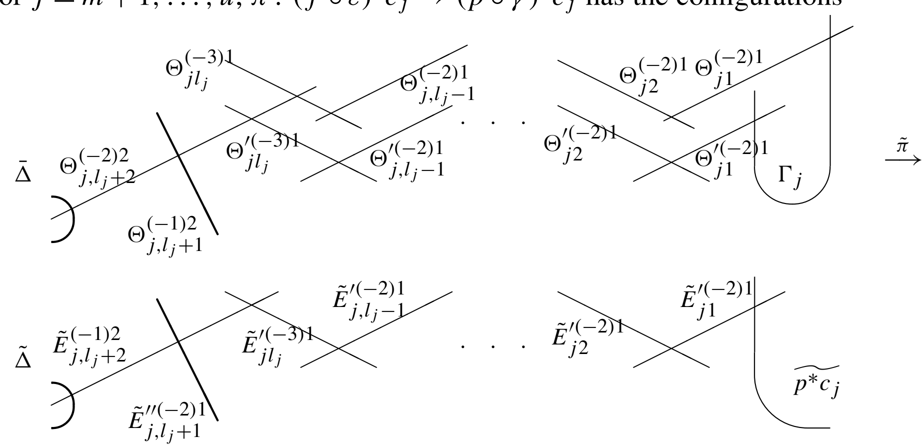


where $\bar{\Delta}$ is the inverse image of $\tilde{\Delta}$, thick lines mean ramification or branch locus of $\tilde{\pi}$, and superscript numbers without brackets are multiplicities and superscript numbers within brackets denote self-intersections. Hence

$$f^*c_j = \Theta'_{j,l_j+1} + \Theta''_{j,l_j+1} + 2\Theta'_{j,l_j+2} + 2\sum_{l=1}^{l_j} \Theta_{jl} + 2\sum_{l=1}^{l_j} \Theta'_{jl} + 2\Gamma_j$$

is as in (b_{2l_j+1}) of [Cai 2006a, 2.6]. Θ_{jl} $(l = 1, \ldots, l_j)$ are λ-fixed -2-curves and Γ_j is an λ-fixed elliptic curve.

For $j = m+1, \ldots, u$, $\tilde{\pi} : (f \circ \varepsilon)^* c_j \to (p \circ \gamma)^* c_j$ has the configurations

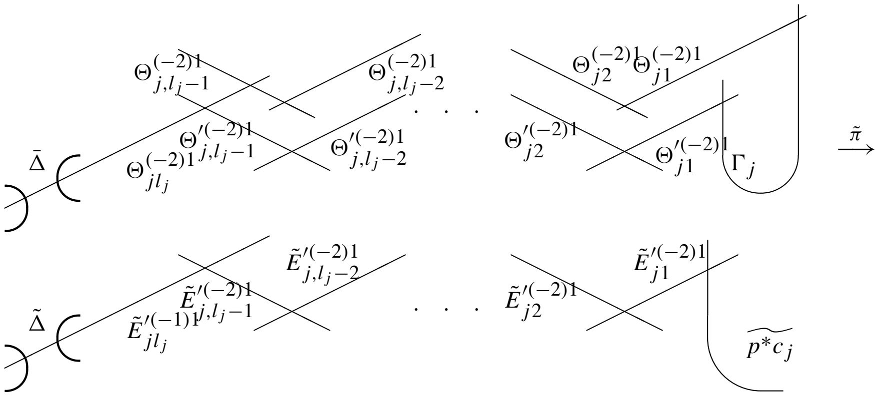


Since $\tilde{\delta}_{|\widetilde{p^*c_j}} = \tilde{C}_{1|\widetilde{p^*c_j}} - \tilde{E}'_{j1|\widetilde{p^*c_j}}$ $(= \tilde{C}_{1|\widetilde{p^*c_j}} - \tilde{E}_{j,l_j+2|\widetilde{p^*c_j}}$ when $l_j = 0)$ is nontrivial, the inverse image Γ_j of $\widetilde{p^*c_j}$ is connected. Hence

$$f^*c_j = \sum_{l=1}^{l_j} \Theta_{jl} + \sum_{l=1}^{l_j} \Theta'_{jl} + \Gamma_j$$

(here we also denote by Θ_{jl_j} and Θ'_{jl_j} the image of Θ_{jl_j} and Θ'_{jl_j} in S) is as in (v) of [Cai 2006a, Lemma 4.9]. The chain of -2-curves in f^*c_j is of type A_{2l_j} and $\Theta_{jl_j} \cap \Theta'_{jl_j}$ is a nonisolated λ-fixed point. (When $l_j = 0$ f^*c_j is an irreducible curve with exactly one node p_j, which is a nonisolated λ-fixed point. The normalization of f^*a_j is an elliptic curve.)

For $j = u+1, \dots, u+v$, $\tilde{\pi} : (f \circ \varepsilon)^* c_j \to (p \circ \gamma)^* c_j$ has the configurations



Since $\tilde{\delta}_{|\widetilde{p^*c_j}} = \tilde{C}_{1|\widetilde{p^*c_j}} - \tilde{E}'_{j1|\widetilde{p^*c_j}}$ is nontrivial, the inverse image Γ_j of $\widetilde{p^*c_j}$ is connected. Hence

$$f^*c_j = \sum_{l=1}^{l_j} \Theta_{jl} + \sum_{l=1}^{l_j-1} \Theta'_{jl} + \Gamma_j$$

is as in (v) of [Cai 2006a, Lemma 4.9]. The chain of -2-curves in f^*c_j is of type A_{2l_j-1}.

For $j = 1, \dots, m$, $\lambda_{|\Theta'_{j,l_j+2}}$ is an involution with fixed points

$$q_j = \Theta_{j,l_j} \cap \Theta'_{j,l_j+2}, \quad q'_j = \bar{\Delta} \cap \Theta_{j,l_j+2}$$

(the former equals $\Gamma_j \cap \Theta'_{j,l_j+2}$ when $l_j = 0$). See the picture above. Since $\bar{\Delta}$ is τ-invariant, q'_j is τ-fixed. From

$$\varepsilon^* K_S = (\gamma \circ \tilde{\pi})^* (p^*(\eta + D) + C_1), \tag{3.0.2}$$

we see that

$$(l_j+1)(\Theta'_{j,l_j+1} + \Theta''_{j,l_j+1}) + (2l_j+1)\Theta'_{j,l_j+2} + \sum_{l=1}^{l_j} 2l\Theta_{jl} + \sum_{l=1}^{l_j} (2l-1)\Theta'_{jl}$$

is contained in the fixed part of $|K_S|$. By [Cai 2006a, 2.9], f^*c_j is not of type V in the sense of Horikawa. So by [Cai 2006a, 2.8], q_j, q'_j are isolated τ-fixed points and there are three nonisolated τ-fixed points r_{1j}, r_{2j}, r_{3j} on Γ_j. So Θ'_{j,l_j+2} is σ-fixed (otherwise, $\langle \lambda, \tau \rangle \hookrightarrow \operatorname{Aut} \Theta'_{j,l_j+2}$, which is a contradiction since $\langle \lambda, \tau \rangle$ is not cyclic) and r_{1j}, r_{2j}, r_{3j} are σ-fixed points. Similarly we see easily that Θ'_{jl} $(l = 1, \dots, l_j)$ are σ-fixed. Hence

$$e((f^*c_j)^\sigma) = 2(l_j+1) + 3 = 2l_j + 5 \quad \text{for } j = 1, \dots, m.$$

For $j = m+1, \ldots, u+v$, since f^*c_j is reduced, by [Cai 2006a, 2.4], σ has no fixed curves on f^*c_j. Since each component of f^*c_j is σ-invariant, each node point of f^*c_j is σ-fixed. We show that they are isolated σ-fixed points. If there is a σ-fixed point $x \in f^*c_j$ which is not isolated, then there is a σ-fixed curve D (necessarily being horizontal with respective to f) passing through x. Since D is contained in the fixed part of $|K_S|$, $Df^*c_j = 2$. This implies there are three σ-invariant curves meeting in x with distinct tangent directions, and hence the induced linear action of σ on the tangent space at x must be ςid for some $\varsigma \in \mathbb{C}$, a contradiction. (When $m+1 \le j \le u$ and $l_j = 0$, then p_j is a nonisolated τ-fixed point by [Cai 2001, Lemma 2.4], both τ and λ exchange the local branches at p_j. So σ fixes the local branches at p_j, implying that p_j is an isolated fixed point of σ.) Hence

$$e((f^*c_j)^\sigma) = \begin{cases} 2l_j + 1, & j = m+1, \ldots, u; \\ 2l_j, & j = u+1, \ldots, u+v. \end{cases}$$

Let $H \subset S$ be the inverse image of C_1. Both $\tau_{|H}$ and $\lambda_{|H}$ are involution of H. (Clearly by (3.0.2), H is contained in the fixed part of $|K_S|$. So H is τ-invariant and $H_{|F}$ is a g_2^1 on F, where F is a general fiber of f. If $\tau_{|H}$ = identity, let $H \cap F = \{s, s'\}$, then $s + s' = H_{|F} \equiv 2s$, which implies $s' \equiv s$ on F, a contradiction.) So H is a σ-fixed curve. Clearly H is the only σ-fixed curve which is horizontal with respective to f. we show that $f_{|H} : H \to C$ is étale. In particular, this implies r_{1j}, r_{2j}, r_{3j} are isolated σ-fixed points. Suppose $x \in H$ is a ramification point of $f_{|H}$. Let $F' = f^*(f(x))$. Since $HF' = 2$, we have $H \cap F' = \{x\}$. Since H is λ-invariant, we have x is $\langle \tau, \lambda \rangle$ -fixed. Since $\langle \tau, \lambda \rangle$ is not cyclic, x is a singular point of F'. If $F' = f^*c_j$ for some j, $m+1 \le j \le u+v$, then x is one of the node points of f^*c_j, which is a contradiction since these points are isolated fixed points of σ. Now we suppose $F' = f^*c_j$ for some j, $1 \le j \le m$. Since Θ'_{j,l_j+2} is σ-fixed, Θ'_{j,l_j+1} is not σ-fixed. So there is a σ-fixed point o_j on $\Theta'_{j,l_j+1} \setminus \Theta'_{j,l_j+1} \cap \Theta'_{j,l_j+2}$. By [Cai 2001, Lemma 2.4], H passes through o_j, which is a contradiction. Since H is étale over C, $e(H) = 0$. Summing-up, we have

$$\begin{aligned} e(S^\sigma) &= \sum_{j=1}^{u+v} e((f^*c_j)^\sigma) + e(H) = \sum_{j=1}^{m} (2l_j + 5) + \sum_{j=m+1}^{u} (2l_j + 1) + \sum_{j=u+1}^{u+v} 2l_j \\ &= 2 \sum_{j=1}^{u+v} l_j + 4m + u. \end{aligned}$$

By the Noether formula, $e(S) = 8n$. Applying the topological Lefschetz formula to σ [Atiyah and Singer 1968, p. 566], namely

$$e(S) + 8(q(S) - \dim H^0(S, \Omega^1_S)^\sigma) - 2(H^2(S, \mathbb{Q}) - \dim H^2(S, \mathbb{Q})^\sigma) = e(S^\sigma),$$

we get

$$2(\dim H^2(S,\mathbb{Q})-\dim H^2(S,\mathbb{Q})^\sigma)=8n-\Big(2\sum_{j=1}^{u+v}l_j+4m+u\Big)=0$$

by (3.0.1). Thus σ acts trivially on $H^2(S,\mathbb{Q})$, and Lemma 2.8 is proved. □

Remark 3.1. Here is a sketch of an alternative proof of Lemma 2.8 suggested by the referee if T is as in Proposition 2.2(i). In this case $q(S)=g(C)$, and we can use Theorem 3 of [Shioda 1999] to compute the rank of the Néron–Severi group $\mathrm{NS}(S)_{\mathbb{Q}}=\mathrm{NS}(S)\otimes\mathbb{Q}$ of S. Consequently, $\mathrm{NS}(S)_{\mathbb{Q}}$ is generated by H, F and all irreducible components of singular fibers of f. By the construction of S, we can check that H, F and each such component are σ-invariant. Hence σ acts trivially on $\mathrm{NS}(S)_{\mathbb{Q}}$. Let $\mathrm{T}(S)$ be the orthogonal complement of $\mathrm{NS}(S)_{\mathbb{Q}}$ in $H^2(S,\mathbb{Q})$. Note that $\mathrm{T}(S)$ is the smallest rational subspace of $H^2(S,\mathbb{Q})$ such that the complexification of $\mathrm{T}(S)$ contains $H^{2,0}(S)$. Since the involution σ acts trivially on $H^0(\omega_S)$, we have $\mathrm{T}(S)^\sigma=\mathrm{T}(S)$. Hence σ acts trivially on $H^2(S,\mathbb{Q})$.

4. Surfaces with $K_S^2=8\chi(\mathbb{O}_S)$

In this section, we describe explicitly families of pairs (S,σ), where S is a fiber surface of genus 2 with $K_S^2=8\chi(\mathbb{O}_S)$, and σ is an involution of S inducing trivial action on $H^2(S,\mathbb{Q})$.

Throughout the section, we denote by τ_D the hyperelliptic involution of a hyperelliptic curve D; for a point e in an elliptic curve E, we denote by t_e the translation by e.

Example 4.1. Let $(S,\sigma)=(F\times C,\ \tau_F\times\tau_C)$, where F and C are hyperelliptic curves with $g(F)=2$ and $g(C)\geq 2$. This example is well known.

Example 4.2. Let F be a curve of genus 2 with a bielliptic involution λ_F. Let $\tilde{B}=\mathbb{P}^1$ and $\gamma_{\tilde{B}}$ an involution of $\tilde{B}$. Let $\pi: C\to B:=\tilde{B}/\langle\gamma_{\tilde{B}}\rangle$ be a double cover with $g(C)\geq 2$, such that the branch points of $\tilde{B}\to\tilde{B}/\langle\gamma_{\tilde{B}}\rangle=B$ are contained in that of π. Let $\tilde{C}$ be the normalization of $C\times_B\tilde{B}$ and $\gamma_{\tilde{C}}\in\mathrm{Aut}\tilde{C}$ the lift of $\gamma_{\tilde{B}}$. (Note that $\tilde{C}$ is hyperelliptic since the involution corresponding to $\tilde{C}\to\tilde{B}$ is the hyperelliptic one.)

Let $(S,\sigma)=((F\times\tilde{C})/\langle\lambda_F\times\gamma_{\tilde{C}}\rangle,\ \overline{\tau_F\times\tau_{\tilde{C}}})$, where $\overline{\tau_F\times\tau_{\tilde{C}}}$ is the involution of $(F\times\tilde{C})/\langle\lambda_F\times\gamma_{\tilde{C}}\rangle$ induced by $\tau_F\times\tau_{\tilde{C}}$.

Example 4.3. Let G be one of the groups $\mathbb{Z}_a$ $(a=2,3,4,5,6,8,10)$ or $\mathbb{Z}_b\oplus\mathbb{Z}_2$ $(b=2,6)$. Let F be a curve of genus 2 on which G acts faithfully and $g(F/G)=0$. Let $\tilde{B}$ be an elliptic curve and G a subgroup of translations of $\tilde{B}$. Let $C\to B:=\tilde{B}/G$ be a double cover with $g(C)\geq 2$. Let $\tilde{C}=C\times_B\tilde{B}$. Then G induces a faithful action on $\tilde{C}$. Let $\lambda_{\tilde{C}}$ be the involution of $\tilde{C}$ corresponding to the double cover $\tilde{C} \to \tilde{B}$.

Let $(S, \sigma) = ((F \times \tilde{C})/G,\ \overline{\tau_F \times \lambda_{\tilde{C}}})$, where G acts on $F \times \tilde{C}$ via a product action.

Example 4.4. Let F be a curve of genus 2 with a bielliptic involution λ_F. Let $\tilde{B}$ be an elliptic curve, and $e \in \tilde{B}$ a nontrivial 2-torsion point. Let $\pi : C \to B := \tilde{B}/\langle t_e, -1_{\tilde{B}}\rangle$ be a double cover such that the branch locus of $\tilde{B} \to B$ is contained in that of π. Let $\tilde{C}$ be the normalization of $C \times_B \tilde{B}$, and $\widetilde{-1_{\tilde{B}}}, \tilde{t}_e \in \operatorname{Aut} \tilde{C}$ the lifts of $-1_{\tilde{B}}, t_e \in \operatorname{Aut} \tilde{B}$ respectively. Let $\lambda_{\tilde{C}}$ be the involution of $\tilde{C}$ corresponding to the double cover $\tilde{C} \to \tilde{B}$.

Let $(S, \sigma) = ((F \times \tilde{C})/\langle \tau_F \times \tilde{t}_e,\ \lambda_F \times (\widetilde{-1_{\tilde{B}}})\rangle,\ \overline{\tau_F \times \lambda_{\tilde{C}}})$.

Example 4.5. Let $\tilde{B}$ be an elliptic curve, and $e \in \tilde{B}$ a nontrivial 4-torsion point. Let $G := \langle t_e, -1_{\tilde{B}}\rangle \simeq D_8$ (the dihedral group of order 8). Let F be a curve of genus 2 on which G acts faithfully. Let $\pi : C \to B := \tilde{B}/G$ be a double cover such that the branch locus of $\tilde{B} \to B$ is contained in that of π. Let $\tilde{C}$ be the normalization of $C \times_B \tilde{B}$. Then G induces a faithful action on $\tilde{C}$. Let $\lambda_{\tilde{C}}$ be the involution of $\tilde{C}$ corresponding to the double cover $\tilde{C} \to \tilde{B}$.

Let $(S, \sigma) = ((F \times \tilde{C})/G,\ \overline{\tau_F \times \lambda_{\tilde{C}}})$.

Remark 4.6. Let (S, σ) be as in one of Examples 4.1–4.5. Clearly S has a fibration of genus 2 with $K_S^2 = 8\chi(\mathbb{O}_S)$. Applying the topological and holomorphic Lefschetz formula to σ (see [Atiyah and Singer 1968, p. 566]) or by [Cai 2006b, 3.1], we can check easily that σ induces trivial actions on $H^2(S, \mathbb{Q})$.

Theorem 4.7. *Let S be a complex nonsingular projective surface of general type with $\chi(\mathbb{O}_S) \geq 5$, and $f : S \to C$ be a relatively minimal fibration of genus 2. Let $G \subset \operatorname{Aut} S$ be a nontrivial subgroup of automorphisms of S, inducing trivial actions on $H^2(S, \mathbb{Q})$. Assume that the canonical map ϕ_S of S is generically finite. Then $|G| = 2$, $g(C) \geq 2$, the generator σ of G induces a hyperelliptic involution or a bielliptic involution $\bar{\sigma}$ of C such that $\bar{\sigma} \circ f = f \circ \sigma$, and either*

(4.7.1) $K_S^2 = 8\chi(\mathbb{O}_S)$ *and* $g(C) \leq q(S) \leq g(C) + 2$,

(4.7.2) $K_S^2 = 8\chi(\mathbb{O}_S) - 6$ *and* $g(C) \leq q(S) \leq g(C) + 1$, *or*

(4.7.3) $K_S^2 = 8\chi(\mathbb{O}_S) - 12$ *and* $q(S) = g(C)$.

Moreover, if S is as in (4.7.1), *then* (S, σ) *belongs to one of Examples 4.1–4.5.*

Proof. The first part of this theorem follows from [Cai 2006b, Theorem 1.1]. Now we let $f : S \to C, \sigma$ be as in (4.7.1). Let τ be the hyperelliptic involution of

$f: S \to C$, and $\lambda = \sigma \circ \tau$. We have a commutative diagram

$$\begin{tikzcd} S \arrow[dr, "f"'] & \tilde{S} \arrow[l, "\rho"'] \arrow[r, "\tilde{\pi}"] \arrow[d] & \tilde{S}/\tilde{\lambda} \arrow[r, "\eta"] \arrow[d] & T \arrow[dl, "h"] \\ & C \arrow[r, "\pi"] & B := C/\bar{\sigma} & \end{tikzcd}$$

where ρ is the blowup of all isolated fixed points of λ, $\tilde{\lambda}$ the induced involution on $\tilde{S}$, and η is the blowdown of all -1-curves contained in fibers of $\tilde{S}/\tilde{\lambda} \to B$. Then $p_g(T) = 0$, and $h: T \to B$ is a relatively minimal fibration of genus 2. The configurations of reducible fibers of h is as in Table 1 (see [Cai 2006b, 2.9]), where $q_f = q(S) - g(C)$, and $4(b_0)$, etc (column 5) means h having 4 reducible fibers of type (b_0) and no other reducible fibers.

	q_f	$g(B)$	$q(T)$	K_T^2	configurations of reducible singular fibers of h
1	0	0	0	0	$4(b_0)$
2	0	1	1	0	a nontrivially analytic fiber bundle
3	1	0	1	-4	$2(b_0)$
4	2	0	2	-8	a trivial fiber bundle

Table 1

Since f is a fiber bundle by [Xiao 1985, p. 18], h has constant moduli. Let F be a general fiber of h. There exists a finite group G acting on F and on some smooth curve $\tilde{B}$ such that h is birationally isomorphic to $(F \times \tilde{B})/G \to \tilde{B}/G$.

If h is as in line 4 of Table 1, then clearly (S, σ) is as in Example 4.1.

Case 1: h is as in line 3 of Table 1. In this case $g(F/G) = q(T) - g(\tilde{B}/G) = 1$. So $|G| = 2$ by the Hurwitz formula. Since $p_g(T) = 0$, we have $\tilde{B} \simeq \mathbb{P}^1$. So T is birationally isomorphic to $(F \times \tilde{B})/\langle \lambda_F \times \gamma_{\tilde{B}} \rangle$, where λ_F is a bielliptic involution of F, and $\gamma_{\tilde{B}}$ is an involution of $\tilde{B}$. We have a commutative diagram

$$\begin{tikzcd} \tilde{C} \arrow[r, "\mu"] \arrow[dr] & C \times_B \tilde{B} \arrow[r] \arrow[d] & \tilde{B} \arrow[d] \\ & C \arrow[r, "\pi"] & B \end{tikzcd}$$

where π is as in the beginning of the proof and μ is the normalization. Let $\lambda_{\tilde{C}}$ be the involution of $\tilde{C}$ corresponding to the double cover $\tilde{C} \to \tilde{B}$, and $\gamma_{\tilde{C}} \in \operatorname{Aut} \tilde{C}$ is the lift of $\gamma_{\tilde{B}}$. Since the image of reducible fibers of h is contained in the set of branch points of π, the branch points of $\tilde{B} \to B$ are contained in that of π. This implies $\tilde{C} \to C \simeq \tilde{C}/\langle \gamma_{\tilde{C}} \rangle$ is étale. We have a commutative diagram

$$\begin{tikzcd}
F \times (C \times_B \tilde{B}) \simeq C \times_B (F \times \tilde{B}) \arrow[r] \arrow[d] & F \times \tilde{B} \arrow[d] \\
C \times_B (F \times \tilde{B})/\langle \lambda_F \times \gamma \rangle \arrow[r] & (F \times \tilde{B})/\langle \lambda_F \times \gamma \rangle
\end{tikzcd}$$

Hence $S = (F \times \tilde{C})/\langle \lambda_F \times \gamma_{\tilde{C}} \rangle$ and $\sigma = (\overline{\tau_F \times \mathrm{id}_{\tilde{C}}})(\overline{\mathrm{id}_F \times \tau_{\tilde{C}}}) = \overline{\tau_F \times \tau_{\tilde{C}}}$. So (S, σ) is as in Example 4.2.

Case 2: h is as in line 2 of Table 1. In this case, $T \simeq (F \times \tilde{B})/G$, where F, $\tilde{B}$ and G are as in Example 4.3. (Since G is an abelian subgroup of $\mathrm{Aut} F$, we have $|G| \le 4g(F)+4 = 12$ ($\le 4g(F)+2 = 10$ if G is cyclic). Moreover, when $\tau_F \notin G$, since $\langle \tau_F, G \rangle$ is also abelian, we have $|G| = \frac{1}{2}|\langle \tau_F, G \rangle| \le 2g(F)+2 = 6$. Finally $G \not\simeq \mathbb{Z}_2 \oplus \mathbb{Z}_4$ by the Riemann's existence theorem (see, for instance, [Broughton 1991, Proposition 2.1 or Theorem 4.1]). By the same argument as in Case 1, we get (S, σ) is as in Example 4.3.

Case 3: h is as in line 1 of Table 1. Let $B' \to B$ be the double cover branched at four points, which are the image of four singular fibers of type (b_0) of h. Let $T'' \to T \times_B B'$ be the normalization, and $h' : T' \to B'$ the relatively minimal fibration induced by contracting -1-curves contained in the fibers of $T'' \to B'$. Since h has only 4 reducible fibers of type $(b_0(I_0))$, $(b_0(I_1))$ or $(b_0(II))$ (see [Cai 2006b, Table 1]) and no other reducible fibers, by the construction, each singular fiber (if any) of h' is irreducible and reduced. Since h' has constant moduli, this implies h' is a fiber bundle. By [Cai 2006b, Lemma 2.5], $q(T') = 1$. So $(h' : T' \to B') \simeq ((F \times \tilde{B})/G \to \tilde{B}/G)$, where F, $\tilde{B}$ and G are as in Example 4.3. This implies h has only 4 reducible fibers of type $(b_0(I_0))$ and no other singular fibers. Hence, for any $z \in \tilde{B}$, the order of the stabilizer G_z of z in G is at most 2, and if G_z is not trivial for some $z \in \tilde{B}$, then $|G|/|G_z| \le 4$ and the generator of G_z acts on F as a bielliptic involution. So $|G| = 4$ or 8. If $|G| = 4$, then $G \simeq \mathbb{Z}_2^2$ and T is birationally isomorphic to $(F \times \tilde{B})/\langle \tau_F \times t_e,\ \lambda_F \times (-1_{\tilde{B}}) \rangle$, where $e \in \tilde{B}$ is a nontrivial 2-torsion point, and λ_F is the involution of F corresponding to the generator of G_z. If $|G| = 8$, then $G \simeq \mathbb{Z}_8$, Q_8 or D_8 by [Broughton 1991, Theorem 4.1]. Since $G \hookrightarrow \mathrm{Aut}\tilde{B}$, $G \simeq G_1 \rtimes G_2$ (a semidirect product), where G_1 is a group of translations and $G_2 \subset \mathrm{Aut}\tilde{B}$ is a subgroup preserving the group structure. Since $\tilde{B}/G = \mathbb{P}^1$, $G_2 \neq 0$, thus $G_2 \simeq \mathbb{Z}_m$ ($m = 2$ or 4). This implies $G \not\simeq \mathbb{Z}_8$ or Q_8. Hence $G \simeq \langle t_e,\ -1_{\tilde{B}} \rangle \simeq D_8$, where $e \in \tilde{B}$ is a nontrivial 4-torsion point. Now by the similar argument as in Case 1, we get (S, σ) is as in Examples 4.4 and 4.5. □

Remark 4.8. Let S be a surface isogenous to a product of curves of genus at least 2 (see [Catanese 2000; 2003] for properties of these surfaces), and $G \subset \mathrm{Aut}\, S$ be a *nontrivial* subgroup of automorphisms of S, inducing trivial actions on $H^2(S, \mathbb{Q})$.

It is interesting to classify pairs (S, G). Note that fiber surfaces of genus 2 with $K_S^2 = 8\chi(\mathcal{O}_S)$ are isogenous to products of curves. Theorem 4.7 gives a classification for such pairs under the condition that one curve of the products has genus 2.

Acknowledgments

This work was done during the author's stay at the Universität Bayreuth. He is grateful to Professor Fabrizio Catanese for his invitation and for invaluable discussions and help. He also thanks the numbers of Lehrstuhl Mathematik VIII, the Department of Mathematics, in particular Professors Ingrid Bauer, Thomas Peternell, for their help and hospitality. This work has been supported by the DFG-NSFC Chinese-German project "Komplexe Geometrie" and has been also partially supported by the NSFC (No. 10671003). The author is grateful to the referee for several valuable suggestions.

References

[Atiyah 1958] M. F. Atiyah, "On analytic surfaces with double points", *Proc. Roy. Soc. London. Ser. A* **247** (1958), 237–244. MR 20 #2472 Zbl 0135.21301

[Atiyah and Singer 1968] M. F. Atiyah and I. M. Singer, "The index of elliptic operators. III", *Ann. of Math.* (2) **87** (1968), 546–604. MR 38 #5245 Zbl 0164.24301

[Barth et al. 1984] W. Barth, C. Peters, and A. Van de Ven, *Compact complex surfaces*, Ergebnisse der Mathematik **4**, Springer, Berlin, 1984. MR 86c:32026 Zbl 0718.14023

[Beauville 1983] A. Beauville, *Complex algebraic surfaces*, London Mathematical Society Lecture Note Series **68**, Cambridge Univ. Press, Cambridge, 1983. MR 85a:14024 Zbl 0512.14020

[Broughton 1991] S. A. Broughton, "Classifying finite group actions on surfaces of low genus", *J. Pure Appl. Algebra* **69**:3 (1991), 233–270. MR 92b:57021 Zbl 0722.57005

[Burns and Rapoport 1975] D. Burns, Jr. and M. Rapoport, "On the Torelli problem for kählerian $K-3$ surfaces", *Ann. Sci. École Norm. Sup.* (4) **8**:2 (1975), 235–273. MR 56 #5945 Zbl 0324.14008

[Cai 2001] J.-X. Cai, "On abelian automorphism groups of fibred surfaces of small genus", *Math. Proc. Cambridge Philos. Soc.* **130**:1 (2001), 161–174. MR 2002a:14049 Zbl 0974.14028

[Cai 2004] J.-X. Cai, "Automorphisms of a surface of general type acting trivially in cohomology", *Tohoku Math. J.* (2) **56**:3 (2004), 341–355. MR 2005f:14078 Zbl 1076.14052

[Cai 2006a] J.-X. Cai, "Automorphisms of fiber surfaces of genus 2, inducing the identity in cohomology", *Trans. Amer. Math. Soc.* **358**:3 (2006), 1187–1201. MR 2007b:14090 Zbl 1087.14029

[Cai 2006b] J.-X. Cai, "Automorphisms of fiber surfaces of genus 2, inducing the identity in cohomology. II", *Internat. J. Math.* **17**:2 (2006), 183–193. MR 2007b:14091 Zbl 05018011

[Catanese 2000] F. Catanese, "Fibred surfaces, varieties isogenous to a product and related moduli spaces", *Amer. J. Math.* **122**:1 (2000), 1–44. MR 2001i:14048 Zbl 0983.14013

[Catanese 2003] F. Catanese, "Moduli spaces of surfaces and real structures", *Ann. of Math.* (2) **158**:2 (2003), 577–592. MR 2005d:14053 Zbl 1042.14011

[Cornalba 1981] M. Cornalba, "Una osservazione sulla topologia dei rivestimenti ciclici di varietà algebriche", *Boll. Un. Math. Ital. A* (5) **18** (1981), 323–328. Zbl 0462.14007

[Hartshorne 1977] R. Hartshorne, *Algebraic geometry*, Graduate Texts in Mathematics **52**, Springer, New York, 1977. MR 57 #3116

[Iitaka 1971] S. Iitaka, “Deformations of compact complex surfaces, III”, *J. Math. Soc. Japan* **23** (1971), 692–705. MR 44 #7598

[Persson 1978] U. Persson, “Double coverings and surfaces of general type”, pp. 168–195 in *Algebraic geometry* (Tromsø, 1977), edited by L. D. Olson, Lecture Notes in Math. **687**, Springer, Berlin, 1978. MR 80h:14017 Zbl 0396.14003

[Peters 1979] C. A. M. Peters, “Holomorphic automorphisms of compact Kähler surfaces and their induced actions in cohomology”, *Invent. Math.* **52**:2 (1979), 143–148. MR 80h:32050

[Serrano 1991] F. Serrano, “The Picard group of a quasi-bundle”, *Manuscripta Math.* **73**:1 (1991), 63–82. MR 93b:14020

[Shioda 1999] T. Shioda, “Mordell-Weil lattices for higher genus fibration over a curve”, pp. 359–373 in *New trends in algebraic geometry* (Warwick, 1996), edited by K. Hulek et al., London Math. Soc. Lecture Note Ser. **264**, Cambridge Univ. Press, Cambridge, 1999. MR 2000g:11053 Zbl 0947.14012

[Suwa 1969] T. Suwa, “On hyperelliptic surfaces”, *J. Fac. Sci. Univ. Tokyo Sect. I* **16** (1969), 469–476. MR 42 #2035

[Ueno 1976] K. Ueno, “A remark on automorphisms of Enriques surfaces”, *J. Fac. Sci. Univ. Tokyo Sect. I A Math.* **23**:1 (1976), 149–165. MR 53 #8071

[Xiao 1985] G. Xiao, *Surfaces fibrées en courbes de genre deux*, Lecture Notes in Mathematics **1137**, Springer, Berlin, 1985. MR 88a:14042

Received April 5, 2006. Revised February 1, 2007.

JIN-XING CAI
LMAM, SCHOOL OF MATHEMATICAL SCIENCES
PEKING UNIVERSITY
BEIJING 100871
CHINA

jxcai@math.pku.edu.cn

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

VECTOR FIELDS, TORUS ACTIONS AND EQUIVARIANT COHOMOLOGY

JIM CARRELL, KIUMARS KAVEH AND VOLKER PUPPE

We give an explicit connection between the holomorphic equivariant cohomology as defined by Carrell and Lieberman and the usual equivariant cohomology of Borel and Cartan.

Let X be a smooth complex projective variety equipped with a $\mathbb{C}^*$-action with fixed point set Z. By results of Carrell and Lieberman, there exists a filtration $F_0 \subset F_1 \subset \cdots$ of $H^*(Z, \mathbb{C})$ such that $\mathrm{Gr}H^*(Z, \mathbb{C}) \cong H^*(X, \mathbb{C})$ as graded algebras. We give here an explicit connection between this filtration and the $\mathbb{C}^*$-equivariant cohomology of X.

1. Introduction

Let X denote a compact Kähler manifold, and suppose V denotes a holomorphic vector field on X whose zero set Z is nonempty. Let Ω_X^p denote the sheaf of holomorphic p-forms on X. The contraction operator i_V defines a complex of sheaves

$$0 \to \Omega_X^n \to \Omega_X^{n-1} \to \cdots \to \Omega_X^1 \to \Omega_X^0 \to 0,$$

where $n = \dim X$, and an old result of the first author and David Lieberman [1973; 1977] states that the spectral sequence associated to this complex degenerates at its E_1 term, namely $H^*(X, \Omega^*)$ (see Section 2 for a review of this result). This fact, which uses the Deligne Degeneracy Criterion, implies the vanishing statement

$$H^p(X, \Omega^q) = 0 \quad \text{if } |p-q| < \dim Z,$$

and yields a description of the Dolbeault cohomology algebra $H^*(X, \Omega_X^*)$ of X as the graded $\mathbb{C}$-algebra associated to the filtration of the hypercohomology $H^*(K_X)$, which is a ring since i_V is a derivation. Although this result has enabled descriptions of cohomology in a number of special cases, for example, algebraic homogeneous spaces [Akyıldız 1982; Carrell 1992], Schubert varieties [Carrell 1992] and toric varieties [Kaveh 2005], the proof itself in [Carrell and Lieberman 1977] doesn't give any insight into how the filtration can be described. This problem is the motivation of the present paper. In fact, we will show that equivariant cohomology

MSC2000: primary 14F43; secondary 57R91.
Keywords: equivariant cohomology, holomorphic vector field, equivariant vector field, torus action.

and localization give a more transparent way of approaching the theory, provided that V is generated by a $\mathbb{C}^*$-action, which also solves the filtration question.

Throughout the paper, V will denote a holomorphic vector field generated by a $\mathbb{C}^*$-action. The only assumption on the fixed point set Z of this action is that it be nonempty. It is well known that Z is also a smooth Kähler subvariety. Let $H^{p,q}(X) = H^q(X, \Omega^p)$, and recall the Hodge decomposition of the cohomology algebra of X:

$$H^*(X, \mathbb{C}) = \bigoplus_{p+q=*} H^{p,q}(X).$$

Also, for each $s \in \mathbb{Z}$, put

$$\mathcal{H}^s(X) = \bigoplus_{q-p=s} H^{p,q}(X).$$

Then $\mathcal{H}^*(X)$ is a graded $\mathbb{C}$-algebra. Note that $H^*(X, \mathbb{C}) = \bigoplus_s \mathcal{H}^s(X)$ (but not as graded algebras). The following result summarizes what is known in this setting.

Theorem 1.1. *Let X be compact Kähler and admit a $\mathbb{C}^*$-action with a nonempty fixed point set Z. Let V be the holomorphic vector field on X determined by this action and K_X^* the hypercohomology determined by the spectral sequence associated to V. Then, for all $s \in \mathbb{Z}$:*

(i) $\dim H^s(K_X^*) = \dim \mathcal{H}^s(X)$;

(ii) *there exists a $\mathbb{C}$-algebra isomorphism*

$$H^s(K_X^*) \cong \mathcal{H}^s(Z);$$

(iii) *we have*

$$\sum_{q-p=s} \dim H^{p,q}(X) = \sum_{q-p=s} \dim H^{p,q}(Z);$$

(iv) *there exists a filtration of $\mathcal{H}^*(Z)$ that yields an isomorphism of graded rings*

$$\bigoplus_s \mathcal{H}^s(X) \cong \operatorname{Gr} \mathcal{H}^*(Z).$$

In the above, (i) follows from the degeneracy of the spectral sequence of V. The isomorphism (ii) is proven in [Carrell and Sommese 1979], and (iii) follows from the first two parts. The last is in fact treated in several papers, for example, [Carrell and Sommese 1979; Fujiki 1979; Ginzburg 1987]. Also see [Feng 2003] for a proof that doesn't use $\mathbb{C}^*$-actions but assumes V vanishes transversely along Z.

Spaces admitting a $\mathbb{C}^*$-action often have the property that $H^{p,q}(X) = 0$ if $p \neq q$ (for example, algebraic homogeneous spaces, projective toric varieties and, more generally, spherical varieties). For such X, $H^{2p}(X, \mathbb{C}) = H^p(X, \Omega^p)$ and $H^{2p+1}(X, \mathbb{C}) = 0$ for all $p \geq 0$. By (iii), the same is true for Z. Thus, $\mathcal{H}^s(X) = H^{2s}(X, \mathbb{C})$ and similarly $\mathcal{H}^*(Z) = H^*(Z, \mathbb{C})$. Hence the map in (iv) reduces to the graded $\mathbb{C}$-algebra isomorphism

$$H^*(X, \mathbb{C}) \;\cong\; \mathrm{Gr}\, H^*(Z, \mathbb{C}).$$

We note, however, that the filtration on $H^*(Z, \mathbb{C})$ has nothing to do with the natural filtration arising from the usual grading of cohomology.

The plan of the paper is as follows. We will use Section 2 to review the spectral sequence of a holomorphic vector field and Section 3 to recall some basic facts about equivariant cohomology and the Cartan complex. Our main results, Theorems 4.2 and 4.4, are proved in Section 4. Theorem 1.1 follows readily from these two results. In Section 5, we give a simple proof of a result in [Carrell 1995] on regular actions, namely, actions of the 2×2 upper triangular matrices over $\mathbb{C}$ of determinant one such that the unipotent subgroup has a unique fixed point. The equivariant cohomology of these varieties was described in [Brion and Carrell 2004]. In Section 6 we consider some examples.

A few comments about the proofs in Section 4 are in order. Let T denote the compact torus in $\mathbb{C}^*$, and suppose $H_T^*(X, \mathbb{C})$ denotes the T-equivariant cohomology of X over $\mathbb{C}$. One knows $H_T^*(X, \mathbb{C})$ is a free $\mathbb{C}[t]$-module of rank $H^*(X, \mathbb{C})$, so, as a $\mathbb{C}[t]$-module, $H_T^*(X, \mathbb{C}) \cong \mathbb{C}[t] \otimes H^*(X, \mathbb{C})$. Recently, Teleman [2000] and Lillywhite [2003] have defined Dolbeault equivariant cohomology groups $H^{p,q}_{T,\bar\partial}(X)$ for X and showed that $H_T^*(X, \mathbb{C})$ admits the usual Hodge decomposition provided X is compact Kähler. This allows us to define the groups $\mathcal{H}^s_T(X)$ analogous to the groups $\mathcal{H}^s(X)$ defined above. We will show that evaluating polynomials at $t = 1$ gives a map (of $\mathbb{C}$-algebras) $\mathcal{H}^*_T(X, \mathbb{C}) \to H^*(K_X^*)$. (This idea is suggested by a paper of the third author [Puppe 1979/80].) The key result Theorem 1.1(ii) follows from localization in equivariant cohomology. The filtration of $H^*(K_X^*)$ essentially turns out to be the image of a canonical filtration of $\mathcal{H}^*_T(Z, \mathbb{C}) \to H^*(K_Z^*) = \mathcal{H}^*(Z)$ via the above “strange” map.

2. Zeros of holomorphic vector fields and cohomology

The purpose of this section is to review the spectral sequence associated to a holomorphic vector field [Carrell and Lieberman 1973; 1977]. Let X denote a connected compact Kähler manifold of dimension n with sheaf of holomorphic functions $\mathbb{O}_X$ and sheaves Ω_X^p of holomorphic p-forms for $p > 0$. The contraction operator $i_V : \Omega_X^p \to \Omega_X^{p-1}$ defines the Koszul complex

$$0 \to \Omega_X^n \to \Omega_X^{n-1} \to \cdots \to \Omega_X^1 \to \mathbb{O}_X \to 0.$$

In addition, for all $\phi, \omega \in \Omega_X^*$,

$$i_V(\phi \wedge \omega) = i_V\phi \wedge \omega + (-1)^p \phi \wedge i_V\omega$$

if $\phi \in \Omega^p_X$. Let $A^{p,q}(X)$ denote the smooth forms on X of type (p,q). The $\overline{\partial}$ operator $A^{p,q} \to A^{p,q+1}$ anticommutes with i_V, so $(\overline{\partial} - i_V)^2 = 0$. Put

$$K^s_X = \bigoplus_{q-p=s} A^{p,q}, \tag{1}$$

and define $D : K^s_X \to K^{s+1}_X$ to be $\overline{\partial} - i_V$. Then because $D^2 = 0$, we obtain cohomology groups $H^s(K^*_X)$. Moreover, K^*_X is a differential graded algebra under the exterior product, so the cohomology groups form a graded $\mathbb{C}$-algebra $H^*(K^*_X)$. Let $F_\bullet = F_0 \subset F_1 \subset F_2 \subset \cdots \subset F_n$ be the filtration of the double complex $A^{*,*}(X)$, with $F_i = \bigoplus_{r \le i} A^{r,*}(X)$. Since i_V is a derivation, we obtain filtrations $F_\bullet H^s(K^*_X)$ for all s such that

$$F_i H^s(K^*_X) F_j H^t(K^*_X) \subset F_{i+j} H^{s+t}(K^*_X).$$

Now consider the spectral sequence

$$E_1^{-p,q} = H^q(X, \Omega^p_X) \Rightarrow H^{q-p}(K^*_X). \tag{2}$$

The main result is:

Theorem 2.1 [Carrell and Lieberman 1973; 1977]. *If V has zeros, then all differentials in* (2) *are trivial. Consequently $E_1 = E_\infty$, and there are $\mathbb{C}$-linear isomorphisms*

$$H^{p+s}\big(X, \Omega^p_X\big) \cong F_p H^s(K^*_X)/F_{p-1} H^s(K^*_X), \tag{3}$$

for every $p \ge 0$ and s which give an isomorphism of bigraded $\mathbb{C}$-algebras

$$\bigoplus_{p,s} H^{p+s}(X, \Omega^p_X) \cong \bigoplus_{p,s} F_p H^s(K^*_X)/F_{p-1} H^s(K^*_X). \tag{4}$$

3. Remarks on equivariant cohomology

In this section, we will briefly recall the two basic definitions of equivariant cohomology due to Borel and Cartan, and state a recent result of Teleman [2000, Theorem 7.3] and Lillywhite [2003, §5.1] on equivariant Dolbeault cohomology. Suppose G is a compact topological group acting on a space M. It is well known that there exists a contractible space EG with a free G-action. The quotient $BG = EG/G$ is called the classifying space of G. Put

$$M_G = (M \times EG)/G.$$

The *equivariant cohomology* of M over $\mathbb{C}$ is defined to be

$$H^*_G(M) = H^*(M_G, \mathbb{C}).$$

If G is a compact torus, say T, then $H_T^*(\text{point}) = H^*(BT)$ is identified with the polynomial ring $S = \mathbb{C}[\mathrm{Lie}(T)]$, which is graded by assigning degree two to linear forms on $\mathrm{Lie}(T)$. Thus, $H_T^*(M)$ is an S-module (via the natural map $\pi : M_T \to BT$), and one has the following fundamental fact:

Theorem 3.1 (Localization Theorem). *Suppose the compact torus T acts on a space M which admits an equivariant imbedding into a representation of T. Then the kernel as well as the cokernel of the canonical map*

$$i^* : H_T^*(M) \to H_T^*(M^T)$$

induced by the inclusion $i : M^T \hookrightarrow M$ are torsion modules over S. Thus if $H_T^*(M)$ is a free module over S, then i^* is injective. Moreover, i^* becomes an isomorphism after inverting elements of a finitely generated multiplicative subset of the polynomial algebra S.

If $H_T^*(M)$ is a free S-module, then the action of T on M is said to be *equivariantly formal*. Equivalently, M is equivariantly formal if the spectral sequence of the fibration $M_T \to BT$ collapses.

Remark 3.2. By a result of Frankel [1959], a $\mathbb{C}^*$-action with fixed points on a compact Kähler manifold is equivariantly formal for the compact torus $T = S^1 \subset \mathbb{C}^*$. More generally, by a theorem of Kirwan, every Hamiltonian T-action on a compact symplectic manifold is equivariantly formal [1984, Proposition 5.8]. Moreover, the hypotheses of Theorem 3.1 hold in the compact symplectic (in particular, compact Kähler) case. For further examples of equivariantly formal spaces, see [Goresky et al. 1998, §14.1].

To recall Cartan's construction of equivariant cohomology [1951], we will assume the space M is a smooth manifold on which T acts smoothly. Let $\Omega^*(M)$ be the De Rham complex of $\mathbb{C}$-valued forms on M. Define $\Omega_T^*(M)$ to be the complex consisting of all the polynomial maps $f : \mathrm{Lie}(T) \to (\Omega^*(M))^T$. Here the superscript denotes the T-invariants. This is equivalent to defining $\Omega_T^*(M) = (\Omega^*(M) \otimes_{\mathbb{C}} S)^T$. In particular

$$\Omega_T^* := \Omega_T^*(\text{point}) = S^T = S.$$

The grading on $\Omega_T^*(M)$ is defined by $\deg(f) = n + 2p$, if $x \mapsto f(x)$ is of degree p in x and $f(x) \in \Omega^n(M)$. The differential

$$d_T : \Omega_T^*(M) \to \Omega_T^*(M)$$

of this complex is defined by

$$(d_T f)(x) = d(f(x)) - i_{V_x} f(x),$$

where i_{V_x} is the contraction with the vector field V_x on M generated by $x \in \mathrm{Lie}(T)$. Then $d_T \circ d_T = 0$ and d_T increases the degree in $\Omega^*(M)$ by 1.

Theorem 3.3 [Cartan 1951]. *$H_T^*(M)$ and $H^*(\Omega_T^*(M), d_T)$ are isomorphic graded $\mathbb{C}$-algebras.*

If M is a complex manifold and T acts via holomorphic transformations, a Dolbeault version of T-equivariant cohomology is constructed in a similar way. For $x \in \mathrm{Lie}(T)$, let $V_x = W_x + \overline{W_x}$ be the splitting of the generating vector field of x into holomorphic and antiholomorphic components. Imitating the Cartan construction, let $A_T^{p,*}(M)$ be the complex of all polynomial maps f from $\mathrm{Lie}(T)$ to $(A^{p,*}(M))^T$. (Note again that this is the same as defining $A_T{}^{p,*}(M) = (A^{p,*}(M) \otimes_{\mathbb{C}} S)^T$). Giving bidegree (1, 1) to the generators of S defines a bigrading on the algebra $A_T^{*,*}(M) = \bigoplus_{p,q} A_T^{p,q}(M)$. Define the differential $\overline{\partial}_T$ on $A_T^{p,*}(M)$ by

$$(\overline{\partial}_T f)(x) = \overline{\partial}(f(x)) - i_{W_x} f(x).$$

The q-th cohomology of the complex $(A_T^{p,*}(M), \overline{\partial}_T)$ is called the (p,q)-th *equivariant Dolbeault cohomology* of M. It is denoted by $H_T^{p,q}(M)$. Finally, put

$$H^m_{T,\overline{\partial}}(M) = \bigoplus_{p+q=m} H_T^{p,q}(M).$$

We now state a recent result of Lillywhite [2003] and Teleman [2000].

Theorem 3.4 (Equivariant Hodge Decomposition). *If X is a compact Kähler manifold with an equivariantly formal T-action by holomorphic transformations, then $H^*_{T,\overline{\partial}}(X)$ is a free S-module, and there exists an isomorphism*

$$H_T^*(X) \cong H^*_{T,\overline{\partial}}(X)$$

of graded $\mathbb{C}$-algebras.

Finally, we recall the definition of the equivariant Chern classes of a vector bundle. Let E be a complex vector bundle over the a space M on which T acts, and suppose E has a linear action of T lifting the action of T. The projection map $p : E \to M$ defines a map from $E_T = E \times_T ET$ to $M_T = X \times_T ET$. This makes E_T a vector bundle over M_T. The r-th equivariant Chern class of E, denoted by $c_r^T(E)$, is defined to be the r-th Chern class of E_T. It is clear that $c_r^T(E) \in H_T^{2r}(M)$.

Remark 3.5. We will need the following fact in Section 5: suppose M is connected and the action of T on M is trivial. Let E be a line bundle with a T-action as above. Let the weight of action of T on each fibre of E be ω. Then

$$c_1^T(E) = -\omega + c_1(E) \tag{5}$$

in $H_T^2(X) = (S \otimes H^*(X))_2$.

4. The main results

Now let X denote a connected compact Kähler manifold of dimension n having a $\mathbb{C}^*$ action with nonempty fixed point set Z, and let T be the compact torus in $\mathbb{C}^*$. Let V be the generating vector field of $1 \in \mathbb{C} = \mathrm{Lie}(\mathbb{C}^*)$, and, as before, let K_X^* denote the total complex of the Koszul complex of the vector field V. It is well known that $X^T = Z$. From now on, $S = \mathbb{C}[t]$.

The purpose of this section is to derive the results about the spectral sequence of V (in particular, to prove Theorem 1.1) using Dolbeault T-equivariant cohomology and to obtain a new picture of the filtration $F_\bullet = F_0 \subset F_1 \subset \cdots \subset F_n$ of $H^*(K_X^*)$.

We first define a chain map $\tilde{\Phi}_X : A_T^{*,*}(X) \to K_X^*$. Recall that an element of $A_T^{*,*}(X)$ is a polynomial map $f : \mathfrak{t} \to (A^{*,*}(X))^T$. By Definition (1), if $f \in A_T^{p,q}(X)$, then $f(1) \in K_X^{q-p}$. Therefore, put

$$\tilde{\Phi}(f) = f(1).$$

Proposition 4.1. *$\tilde{\Phi}$ is a cochain map. That is, for $f \in A_T^{*,*}(X)$, we have*

$$\tilde{\Phi}(\overline{\partial}_T f) = D(\tilde{\Phi}(f)).$$

Proof. $\tilde{\Phi}(\overline{\partial}_T f) = \tilde{\Phi}(\overline{\partial} f(x) - i_{V_x} f(x)) = \overline{\partial} f(1) - i_V f(1) = D(f(1)) = D(\tilde{\Phi}(f))$. Here V_x and V are the generating vector fields of $x \in \mathrm{Lie}(T)$ and $1 \in \mathrm{Lie}(T)$. $\square$

It is now convenient to put $\mathcal{H}_T^s(X) = \bigoplus_i H_T^{i,i+s}(X)$. Note that by Theorem 3.4, $H_T^*(X) = \bigoplus_s \mathcal{H}_T^s(X)$. This gives a new grading on $H_T^*(X)$ by S-submodules. We will denote $H_T^*(X)$ with this grading by $\mathcal{H}_T^*(X)$. By the above proposition, $\tilde{\Phi}$ induces a map

$$\Phi_{X,s} : \mathcal{H}_T^s(X) \to H^s(K_X^*). \tag{6}$$

It is not hard to check that the $\Phi_{X,s}$ give a $\mathbb{C}$-algebra homomorphism.

Let π denote the natural map $\pi : H_T^{p,p+s}(X) \to H^{p,p+s}(X)$ induced by the inclusion $X \hookrightarrow X_T$. By equivariant formality, the ordinary cohomology sequence

$$0 \to S^+ H_T^*(X) \to H_T^*(X) \to H^*(X) \to 0 \tag{7}$$

is exact (compare [Brion 1998, Section 1]), so by the equivariant Hodge decomposition, π is surjective for all p, s.

Let $\mathcal{H}^*(X)$ denote $H^*(X, \mathbb{C})$ and grade it with the decomposition $H^*(X) = \bigoplus_s \mathcal{H}^s(X)$, as defined in Section 2. For any $\mathbb{C}$-vector space V and $a \in \mathbb{C}$, let $V[a]$ denote the S-module structure on V where t acts via multiplication by a. Note that $\dim V[a]$ is the same for all a. By (7), we have another exact sequence of S-modules

$$0 \to S^+ \mathcal{H}_T^s(X) \to \mathcal{H}_T^s(X) \to \mathcal{H}^s(X)[0] \to 0,$$

where S^+ denotes the ideal generated by t. Hence

$$\mathscr{H}^s(X)[0] \cong \mathscr{H}^s_T(X)/S^+\mathscr{H}^s_T(X) \cong \mathscr{H}^s_T(X) \otimes_S \mathbb{C}[0],$$

and therefore

$$\dim(\mathscr{H}^s_T(X) \otimes_S \mathbb{C}[0]) = \dim \mathscr{H}^s(X) = \sum_i \dim H^{i,i+s}(X),$$

We now prove the first three assertions of Theorem 1.1. First, notice that the chain map $\Phi_{X,s}$ in (6) induces a map $\hat{\Phi}_{X,s} : \mathscr{H}^s_T(X) \otimes_S \mathbb{C}[1] \to H^s(K^*_X)[1] = H^s(K^*_X)$.

Theorem 4.2. *The following statements hold for each integer s.*

(i) $\hat{\Phi}_{X,s} : \mathscr{H}^s_T(X) \otimes_S \mathbb{C}[1] \to H^s(K^*_X)$ *is a $\mathbb{C}$-linear isomorphism.*

(ii) *The inclusion mapping $i_Z : Z \to X$ induces a $\mathbb{C}$-algebra isomorphism*

$$i^*_Z : H^s(K^*_X) \cong H^s(K^*_Z) = \mathscr{H}^s(Z).$$

(iii) *In particular*, $\sum_i \dim H^{i,i+s}(X) = \sum_i \dim H^{i,i+s}(Z)$.

Proof. The Localization Theorem 3.1 implies the map i^*_Z induces an isomorphism

$$\mathscr{H}^s_T(X) \otimes_S \mathbb{C}[1] \cong \mathscr{H}^s_T(Z) \otimes_S \mathbb{C}[1].$$

Since $\mathscr{H}^s_T(Z) \otimes_S \mathbb{C}[0] \cong \mathscr{H}^s(Z) = H^s(K^*_Z)$, and since $\dim(\mathscr{H}^s_T(X) \otimes_S \mathbb{C}[a])$ is the rank of $\mathscr{H}^s_T(X)$ as a free S-module for any a, we get an isomorphism $\mathscr{H}^s_T(X) \otimes_S \mathbb{C}[1] \cong H^s(K^*_Z)$, which is nothing more than $i^*_Z \hat{\Phi}_{X,s}$. It follows that $\hat{\Phi}_{X,s}$ is injective.

To prove part (i), it remains to show $\hat{\Phi}_{X,s}$ is surjective. It suffices to show that $\Phi_{X,s}$ is. By standard reasoning about the spectral sequence of a double complex, we have an edge map $e_{p,s} : F_p H^s(K^*_X) \to H^{p,p+s}(X)$ whose kernel contains $F_{p-1} H^s(K^*_X)$. Let $f(t) = \sum_i w_i t^i$, where each $w_i \in A^{p-i,p+s-i}(X)$ represents a class in $H^{p,p+s}_T(X)$. By definition, $\Phi_{X,s}(f) = \sum_i w_i$. Moreover,

$$\pi(f) = w_0 = e_{p,s}\Big(\sum_i w_i\Big).$$

In other words, we get the following commutative diagram.

$$\begin{tikzcd} H^{p,p+s}_T(X) \arrow[r, "\Phi_{X,s}"] \arrow[d, "\pi"] & F_p H^s(K^*_X) \arrow[dl, "e_{p,s}"] \\ H^{p,p+s}(X) & \end{tikzcd}$$

Since π is surjective, it follows from this that $\Phi_{X,s}$ is surjective. This concludes the proof of (i). The statements (ii) and (iii) follow immediately. $\square$

Remark 4.3. Theorem 4.2(i) is analogous to the corollary in [Puppe 1974, p. 13]. The proof of Theorem 4.2 implies that the subcomplex of the Koszul complex consisting of T-invariant forms is quasi-isomorphic to the Koszul complex itself. By first proving this result directly (which is similar to the well-known result that invariant forms in the deRham complex determine the deRham cohomology) and then using that the equivariant Dolbeault evaluated at $t = 1$ is just the invariant Koszul complex, one gets an alternative proof of Theorem 4.2. In this context, the evaluation at $t = 1$ is exact, and hence commutes with homology, whereas the evaluation at $t = 0$ is not.

Theorem 4.2 realizes two of the goals of the paper: a simple proof that i_Z^* is a quasi-isomorphism, and a proof of the isomorphism (3) of Theorem 2.1 that doesn't use the Deligne Degeneracy Criterion. We note that the isomorphism (4) is a formal consequence of the fact that i_V is a derivation.

Let us now comment further on the filtrations. Let $\hat{\Phi}_X : \mathcal{H}_T^*(X) \otimes_S \mathbb{C}[1] \to H^*(K_X^*)$ be the morphism obtained by combining the $\hat{\Phi}_{X,s}$. Note that $\hat{\Phi}_X$ is a $\mathbb{C}$-algebra isomorphism, but not an isomorphism of graded algebras. However, $\mathcal{H}_T^*(X) \otimes_S \mathbb{C}[1]$ and $H^*(K_X^*)$ are both canonically filtered, the former being the filtration induced from the grading on $\mathcal{H}_T^*(X)$ and the latter being the filtration introduced in Section 2. More explicitly, if $p \geq 0$, put $F_p\mathcal{H}_T^s(X) = \bigoplus_{i \leq p} H_T^{i,i+s}(X)$. If $\Phi_{X,s}$ is the map defined in (6), then, by definition,

$$\Phi_{X,s}\big(F_p\mathcal{H}_T^s(X)\big) \subset F_pH^s(K_X^*). \tag{8}$$

Note that $\Phi_{X,s}$ can be described as the map obtained by composing $\hat{\Phi}_{X,s}$ and the natural map from $\mathcal{H}_T^*(X)$ to $\mathcal{H}_T^*(X) \otimes_S \mathbb{C}[1]$ sending α to $\alpha \otimes_S 1$.

We can now give a geometric description of the filtration of $H^*(K_X^*)$. Let $\mathcal{R}_X$ denote the algebra $\mathcal{H}_T^*(X)/S^+\mathcal{H}_T^*(X)$. Since the ideal $S^+\mathcal{H}_T^*(X)$ is homogeneous with respect to the grading of $\mathcal{H}_T^*(X)$, $\mathcal{R}_X$ inherits a grading from $\mathcal{H}_T^*(X)$.

Theorem 4.4. *The mapping Φ_X is a surjection of filtered rings. That is, for all s,*

$$\Phi_{X,s}\big(F_p\mathcal{H}_T^s(X)\big) = F_pH^*(K_X^*),$$

and $\mathcal{R}_X$ is isomorphic to both $\mathcal{H}^*(X)$ and Gr $H^*(K_X^*)$ *as graded algebras.*

Proof. This follows from (8) and Theorem 4.2(i). □

Since the inclusion map $i_Z : Z \to X$ induces a quasi-isomorphism, we immediately obtain a description of the filtration of $H^*(K_Z^*)$ whose associated graded is $\mathcal{H}^*(X)$.

Corollary 4.5. *For each $p \geq 0$,*

$$\Phi_Z \circ i_Z^*\Big(\bigoplus_{0 \leq i \leq p} H_T^{i,i+s}(X)\Big) = F_pH^s(K_Z^*).$$

We will give an example of how to use this result in the next section. Note also that the natural map

$$\Delta_p : H_T^{p,p+s}(X) \to F_p H^s(K_X^*) \to H^p(X, \Omega_X^{p+s})$$

can be described as the p-th derivative map

$$\Delta_p(f) = \frac{1}{p!} f^{(p)}(1).$$

We now use Theorem 4.2 to prove a vanishing theorem which extends the vanishing result $H^{p,q}(X) = 0$ if $|p-q| > \dim Z$.

Theorem 4.6. *If* $|p-q| > \dim Z$, *then* $H_T^{p,q}(X) = 0$.

Proof. Since $H_T^*(Z) = S \otimes_{\mathbb{C}} H^*(Z)$, it follows that

$$H_T^{p,q}(Z) = \bigoplus_{i \le \min(p,q)} S^i \otimes_{\mathbb{C}} H^{p-i,q-i}(Z).$$

But $|p-q| = |(p-i)-(q-i)| > \dim Z$, so $H_T^{p,q}(Z) = 0$ as well. By Theorem 3.4, $H_T^{p,q}(X) \subset H_T^{p+q}(X)$, so the result follows from the Localization Theorem 3.1 since $i^*\big(H_T^{p,q}(X)\big) \subset H_T^{p,q}(Z) \subset H_T^{p+q}(Z)$. □

5. An application

The purpose of this section is to apply our main result to give a simple proof of a fact about the cohomology ring of a regular variety originally proved in [Carrell 1995]. A smooth projective variety X over $\mathbb{C}$ that admits an action of the upper triangular subgroup $\mathfrak{B}$ of $SL_2(\mathbb{C})$ whose unipotent radical $\mathfrak{U}$ has a unique fixed point o is said to be *regular*. Let $\mathfrak{T}$ denote the diagonal torus in $\mathfrak{B}$, and let T be the maximal compact torus in $\mathfrak{T}$. One knows [Carrell 1995] that $X^{\mathfrak{T}} = X^T$ is finite, and, moreover, $o \in X^T$. In fact, let $\lambda : \mathbb{C}^* \to \mathfrak{T}$ be the isomorphism

$$t \to \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}.$$

Then the Bialynicki–Birula cell $X_o = \{x \in X \mid \lim_{t\to\infty} \lambda(t)\cdot x = o\}$ is a $\mathfrak{T}$-invariant open set in X isomorphic with $\mathbb{C}^n$ for $n = \dim X$, and there exist affine coordinates $u_1, \dots, u_n$ on X_o that are quasihomogeneous of positive degree with respect to the $\mathbb{G}_m$-action on X induced by λ. This grading on $\mathbb{C}[X_o] = \mathbb{C}[u_1, \dots, u_n]$ is called the *principal grading*. The *principal filtration* $P_\bullet$ of $\mathbb{C}[X_o]$ is given by

$$P_i \mathbb{C}[X_o] = \sum_{j \le i} \mathbb{C}[X_o]^j,$$

where $\mathbb{C}[X_o]^j$ denotes the subspace generated by the homogeneous elements of degree j. Finally, let

$$\mu(a) = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix},$$

and let T_a ($a \in \mathbb{C}^*$) be the torus $\mu(a)T\mu(a)^{-1}$. We will now prove the following result.

Theorem 5.1. *Suppose X is regular. Then $X^{T_a} \subset X_o$, so $H^0(X^{T_a})$ is a quotient of $\mathbb{C}[X_o]$ for any $a \in \mathbb{C}^*$. Hence it inherits a natural filtration from the principal filtration of $\mathbb{C}[X_o]$, so let $\mathrm{Gr}_P H^0(X^{T_a})$ denote the associated graded ring. Then*

$$H^*(X) \cong \mathrm{Gr}_P H^0(X^{T_a}).$$

Proof. We will only prove the theorem for $a = 1$. The proof for other values of a is similar, after the map Φ_X has been modified. Put $X^T = \{x_1, \ldots, x_r\}$. Now the diagonal action of $\mathfrak{B}$ on $X \times \mathbb{P}^1$ is also regular, with fixed point $(o, 0)$, where 0 represents $[1, 0]$ in $\mathbb{P}^1$. Let

$$Z = \bigcup_{i=1}^{r} \overline{\{(\mu(u) \cdot x_i, u^{-1}) \mid u \neq 0\}}.$$

Let $\mathfrak{X}$ be the reduced intersection $Z \cap (X_o \times \mathbb{C})$. Clearly, $\mathfrak{X}$ is $\mathfrak{T}$-stable, hence its coordinate ring $\mathbb{C}[\mathfrak{X}]$ has a natural (principal) grading. In addition, the projection p_2 induces a $\mathbb{C}[v]$-module structure on the coordinate ring $\mathbb{C}[\mathfrak{X}]$, where v denotes a coordinate function on $\mathbb{C}$.

By a result of Brion and the first author [2004, Theorem 1], the coordinate ring $\mathbb{C}[\mathfrak{X}]$ is isomorphic as a graded $\mathbb{C}$-algebra to the equivariant cohomology algebra $H_T^*(X)$. In fact, an isomorphism

$$\rho : H_T^*(X) \to \mathbb{C}[\mathfrak{X}]$$

is defined as follows. Since the odd cohomology of X is trivial (because X^T is finite), the action $T : X$ is equivariantly formal, so the restriction map $i : H_T^*(X) \to H_T^*(X^T)$ is injective. Note that

$$H_T^*(X^T) = \bigoplus_{i=1}^{r} \mathbb{C}[v]_i,$$

where v is an indeterminate and $\mathbb{C}[v]_i = H_T^*(\{x_i\})$. Thus each $\alpha \in H_T^*(X)$ is determined by an r-tuple of polynomials $(\mathscr{A}_1, \ldots, \mathscr{A}_r)$ in $\mathbb{C}[v]$. Now if $(x, a) \in \mathfrak{X} - (o, 0)$, then $x = \mu(a^{-1}) \cdot x_j$ for a unique index j, where $a \neq 0$. The restriction of α at x_j is a polynomial function $\mathscr{A}_j(v)$. The isomorphism ρ is defined by making $\rho(\alpha)$ the unique function on $\mathfrak{X}$ defined by $\rho(\alpha)(x, v) = \mathscr{A}_j(v)$, if $x = \mu(a^{-1}) \cdot x_j$.

Now note that $H_T^*(X) = \mathscr{H}_T^0(X)$. Furthermore, $\mu(1) : X \to X$ defines an isomorphism $H_T^*(X) \cong H_{T_1}^*(X)$. Thus, we obtain a sequence of maps

$$\mathbb{C}[\mathscr{X}] \xrightarrow{\rho^{-1}} H_T^*(X) \xrightarrow{\mu(1)_*} H_{T_1}^*(X) \xrightarrow{\Phi_X} H^0(K_X^*) = \bigoplus_{X^{T_1}} \mathbb{C},$$

where K_X^* denotes the complex associated to the holomorphic vector field generated by the torus T_1. The composition Ψ_X of these maps sends $F \in \mathbb{C}[\mathscr{X}]$ to the r-tuple $\rho^{-1}(F)(1) = \Phi_X \rho^{-1}(F)$, which, by Theorem 4.4, gives us the result. $\square$

6. Examples

The first example deals with a $\mathbb{G}_m$-action on $\mathbb{P}^n$ having two components of different dimensions.

Example 6.1. Let $X = \mathbb{P}^n$, and let $\mathbb{C}^*$ act on X via

$$t \cdot [a_0, a_1, \ldots, a_n] = [a_0, a_1, \ldots, t a_n].$$

Then $X^T = X_1 \cup X_2$, where $X_1 = \{[0, 0, \ldots, 0, 1]\}$, and $X_2 = V(a_n) \cong \mathbb{P}^{n-1}$. Because $H_T^{p,q}(X) = 0$ for $p \neq q$, we have $H_T^{p,p}(X) = H_T^{2p}(X)$, $\mathscr{H}_T^s(X) = 0$ for $s \neq 0$, and $\mathscr{H}_T^0(X) = H_T^*(X)$. Similarly, $H_T^*(X^T) = \mathscr{H}_T^0(X^T)$. The image of $H_T^*(X)$ in $H_T^*(X^T)$ consists of all triples (α, β, γ) satisfying $\alpha \in H_T^*(X_1) \cong \mathbb{C}[t]$; $\beta \in H_T^*(X_2) \cong \mathbb{C}[t] \otimes H^*(X_2)$ with $\alpha(0) = \beta_0(0)$, where β_0 is the component of β in $\mathbb{C}[t] \otimes H^0(X_2)$; and $\gamma = \sum c_1^T(E_i)$, where the E_i are vector bundles on X_2^T. Recall from (5) that $c_1^T(E) = mt + c_1(E_i)$, where t^m is the weight of the $\mathbb{G}_m$-action on the bundle on X that restricts to E_i. The cochain map Φ_X sends $t \to 1$.

Example 6.2 (Toric varieties [Kaveh 2005]). Let $M = (\mathbb{C}^*)^n$, and let X be a smooth projective M-toric variety. Let $\mathfrak{t} = \mathrm{Lie}(M)$ and $\mathfrak{t}_{\mathbb{R}} \subset \mathfrak{t}$ be the real vector space generated by the lattice of characters of M. Let γ be a 1-parameter subgroup of M in general position in the sense that the fixed point set Z of the $\mathbb{C}^*$-action defined by γ coincides with X^M. Hence $H^{p,q}(X) = 0$ if $p \neq q$, so it follows that $\mathscr{H}^0(X) = H^*(X, \mathbb{C})$. Now let

$$F_0 \subset F_1 \subset \cdots \subset F_n = H^0(Z, \mathbb{C})$$

be the associated filtration. Finally, let Σ be the fan of X in $\mathfrak{t}_{\mathbb{R}}$. Each $z \in Z$ corresponds to a cone of maximal dimension σ_z in Σ.

The equivariant cohomology $H_T^*(X, \mathbb{C})$, where $T = (S^1)^n \subset (\mathbb{C}^*)^n$, can be described as the algebra $\mathscr{A}$ of all continuous functions on $\mathfrak{t}_{\mathbb{R}}$ whose restriction to each cone of Σ is given by a polynomial (conewise polynomial). Under this identification, $H_T^{2i}(X, \mathbb{C})$ corresponds to the subspace $\mathscr{A}_i$ of $\mathscr{A}$ consisting of those functions whose restriction to each cone of maximal dimension is homogeneous of degree i.

Let Q denote the compact torus $\gamma(S^1)$. Then one can verify that the map $\Phi_Z \circ i_Z^*$: $H_Q^*(X) \to H^0(Z, \mathbb{C})$ in Corollary 4.5 sends the restriction to Q of a continuous conewise polynomial function g to an element $\tilde{g} : Z \to \mathbb{C}$ defined by

$$\tilde{g}(z) = g_{|\sigma_z}(\gamma).$$

It follows from Corollary 4.5 that $g \in \mathscr{A}_i$ if and only if $\tilde{g} \in F_i$. The fact that $F_i/F_{i-1} \cong H^{2i}(X, \mathbb{C})$ was verified in [Kaveh 2005] using [Carrell and Lieberman 1977].

Example 6.3 (The flag variety G/B). Let G be a connected semisimple group over $\mathbb{C}$, B a Borel subgroup and $X = G/B$ the flag variety of G. Let H be a maximal (algebraic) torus in B and $\mathfrak{h} = \mathrm{Lie(H)}$. Recall that the fixed point set X^H under left multiplication by H is in one-to-one correspondence with the Weyl group $W = N_G(H)/H$ under the map $w = nH \to nB$. Since $H^{p,q}(X) = 0$ for $p \neq q$, it follows that $H^s(K_X^*) = 0$ if $s \neq 0$ for the holomorphic vector field induced by any one parameter subgroup of H. Now, $H_H^*(X, \mathbb{C})$ is isomorphic as a $\mathbb{C}$-algebra to $S \otimes_{S^W} S$ where S^W denotes the subalgebra of W-invariants (see [Brion 1998, §2 Examples]).

We will first consider the regular case, which is well known but will be used in treating the general case.

(a) Suppose $h \in \mathfrak{h}$ induces a regular one parameter subgroup. That is, $Z = X^H$. Equivalently, the isotropy group W_h of h is trivial. Thus $H^0(K_X^*) = H^0(Z, \mathbb{C}) = \mathbb{C}^W$ under the identification $Z = W$. The map $H_H^*(G/B, \mathbb{C}) \to H^0(Z, \mathbb{C})$ obtained by localizing and setting $t = 1$ is described as follows. Let $S = \mathbb{C}[\mathfrak{h}]$. Now, $H_H^*(X, \mathbb{C})$ is isomorphic as a $\mathbb{C}$-algebra to $S \otimes_{S^W} S$ where S^W denotes the subalgebra of W-invariants (see [Brion 1998, §2 Examples]). Since $H^*(G/B, \mathbb{C})$ is generated by the Chern classes of line bundles, and such line bundles are always H-equivariant, we need only consider the image of an equivariant Chern class $c_1^H(L_\lambda)$, where L_λ denotes the line bundle corresponding to a weight $\lambda \in \mathfrak{h}^*$. But it can be shown that $c_1^H(L_\lambda) = -\sum_{w \in W} 1 \otimes (w \cdot \lambda)$, and so $c_1^H(L_\lambda)$ is sent to the element $f_\lambda \in H^0(Z, \mathbb{C})$ defined by the condition

$$f_\lambda(w) = -\langle w \cdot \lambda, h \rangle. \tag{9}$$

This coincides with the representative of $c_1(L_\lambda)$ on $H^0(Z, \mathbb{C})$ calculated, for example, in [Carrell 1992]. The upshot is that F_1 is the image of $\mathfrak{h}^*$ under the quotient map $S \to \mathbb{C}[W \cdot h]$. This reproves the result that $H^*(X, \mathbb{C}) = \mathrm{Gr}\, \mathbb{C}[W \cdot h]$, where the grading is taken with respect to the filtration obtained as the image of the filtration of S associated to its natural grading. Note that $\mathbb{C}[W \cdot h]$ is the algebra of polynomials on the Weyl group orbit $W \cdot h$.

(b) Suppose the element h is nonregular. Let Φ be the root system of (G, H) and $\Phi_h = \{\alpha \in \Phi \mid \alpha(h) = 0\}$. Put

$$\mathfrak{h}_0 = \bigcap_{\alpha \in \Phi_h} \ker \alpha,$$

and let $H_0 \subset H$ be the corresponding torus. Finally, let L denote the Levi subgroup $L = Z_G(H_0)$. For example, if $G = GL(n, \mathbb{C})$, and H is the diagonal torus, put $h = \mathrm{diag}(a_1 I_{n_1}, a_2 I_{n_2}, \ldots, a_{n_r} I_{n_r})$, where I_l is the $l \times l$ identity matrix, $n_1 + \cdots + n_r = n$ and $a_i \neq a_j$ when $i \neq j$. Then $L = GL(n_1, \mathbb{C}) \times \cdots \times GL(n_r, \mathbb{C})$. Then the Weyl group W_L of L is the isotropy subgroup of h in W. Now $Z = X^{H_0}$ is a union of the flag varieties of L. More precisely, for $w \in W$, let Z_w be the connected component of Z containing $wB \in X^H$. One sees that each Z_w is isomorphic to $L/L \cap B$ and $Z_w = Z_{w'}$ for w, w' in the same right coset of W_L. Thus

$$Z = \bigcup_{w \in W_L \backslash W} Z_w.$$

Hence $H^*(Z, \mathbb{C}) = \bigoplus_{w \in W_L \backslash W} H^*(L/L \cap B)$. To obtain the filtration of $H^*(Z, \mathbb{C})$, take an element $t \in \mathfrak{h}$ that determines a regular 1-parameter subgroup of H. Let $\mathfrak{X} \subset \mathfrak{h} \oplus \mathfrak{h}$ be the W-orbit of (h, t), where W acts diagonally on $\mathfrak{h} \oplus \mathfrak{h}$. One can write

$$\mathfrak{X} = \bigcup_{w \in W_L \backslash W} \mathfrak{X}_w,$$

where $\mathfrak{X}_w = \{(w^{-1} \cdot s, w^{-1} u^{-1} \cdot t) \mid u \in W_L\}$. The elements of $\mathfrak{X}$ are in one-to-one correspondence with X^H, and each $\mathfrak{X}_w$ corresponds to the H-fixed points in Z_w. Let $\mathbb{C}[\mathfrak{X}]$ and $\mathbb{C}[\mathfrak{X}_w]$ denote the coordinate rings of $\mathfrak{X}$ and $\mathfrak{X}_w$, respectively. From part (a), $H^*(Z_w) \cong \mathrm{Gr}\, \mathbb{C}[\mathfrak{X}_w]$ for any $w \in W_L \backslash W$, where the filtration on $\mathbb{C}[\mathfrak{X}_w]$ is induced by the degree. Hence

$$H^*(Z, \mathbb{C}) \cong \bigoplus_{w \in W_L \backslash W} \mathrm{Gr}\, \mathbb{C}[\mathfrak{X}_w].$$

Put $\mathcal{A} = \bigoplus_{w \in W_L \backslash W} \mathrm{Gr}\, \mathbb{C}[\mathfrak{X}_w]$. The following shows that the filtration on $\mathcal{A}$ is induced by the natural filtration on $\mathbb{C}[\mathfrak{X}]$ given by the degree.

Proposition 6.4. *An element $(f_w) \in \mathcal{A}$ lies in F_i if and only if there exists an element $f \in \mathbb{C}[\mathfrak{X}]$ with degree $\leq i$ and whose restriction to $\mathfrak{X}_w$ is a representative for f_w in* $\mathrm{Gr}\, \mathbb{C}[\mathfrak{X}_w]$.

Proof. Note that the result of part (a) implies that $H^*(X, \mathbb{C})$ is generated by $H^2(X, \mathbb{C})$. Hence the filtration is generated by F_1, that is, F_i consists of all polynomials in the elements of F_1 of degree $\leq i$. Hence it is enough to verify the claim for F_1. Consider the line bundle L_λ on X corresponding to a dominant weight λ, and let $L_{\lambda,w}$ be the restriction of L_λ to the small flag variety Z_w. Then, for each $w \in W_L\backslash W$, the weight of the action of s on $L_{\lambda,w}$ is $\langle \lambda, w^{-1}\cdot s\rangle$. From (5),

$$c_1^s(L_{\lambda,w}) = -\langle \lambda, w^{-1}\cdot s\rangle + c_1(L_{\lambda,w}), \tag{10}$$

where c_1^s denotes the equivariant Chern class for the $\mathbb{C}^*$-action induced by s. Then, from Theorem 4.4, (9) and (10) it follows that $c_1(L_{\lambda,w})$ corresponds to the element $(f_{\lambda,w})$ represented by the function

$$(w^{-1}\cdot s, w^{-1}u^{-1}\cdot t) \mapsto -\langle \lambda, w^{-1}\cdot s\rangle - \langle \lambda, w^{-1}u^{-1}\cdot t\rangle. \tag{11}$$

Now let f_λ be the linear function on $\mathfrak{h}\oplus\mathfrak{h}$ given by $f(x, y) = -\lambda(x) - \lambda(y)$. From (11), the restriction of f_λ to $\mathfrak{X}_w$ gives a representative for $f_w \in \operatorname{Gr}\mathbb{C}[\mathfrak{X}_w]$. The Proposition now follows because the $c_1(L_\lambda)$ span $H^2(X, \mathbb{C})$. □

References

[Akyıldız 1982] E. Akyıldız, "Vector fields and cohomology of G/P", pp. 1–9 in *Group actions and vector fields* (Vancouver, 1981), edited by J. B. Carrell, Lecture Notes in Math. **956**, Springer, Berlin, 1982. MR 84j:14054 Zbl 0498.14025

[Brion 1998] M. Brion, "Equivariant cohomology and equivariant intersection theory", pp. 1–37 in *Representation theories and algebraic geometry (Montreal, PQ, 1997)*, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. **514**, Kluwer Acad. Publ., Dordrecht, 1998. Notes by Alvaro Rittatore. MR 99m:14005 Zbl 0946.14008

[Brion and Carrell 2004] M. Brion and J. B. Carrell, "The equivariant cohomology ring of regular varieties", *Michigan Math. J.* **52**:1 (2004), 189–203. MR 2005h:14112 Zbl 1084.14044

[Carrell 1992] J. B. Carrell, "Some remarks on regular Weyl group orbits and the cohomology of Schubert varieties", pp. 33–41 in *Kazhdan-Lusztig theory and related topics (Chicago, IL, 1989)*, Contemp. Math. **139**, Amer. Math. Soc., Providence, RI, 1992. MR 93k:14068 Zbl 0807.14040

[Carrell 1995] J. B. Carrell, "Deformation of the nilpotent zero scheme and the intersection ring of invariant subvarieties", *J. Reine Angew. Math.* **460** (1995), 37–54. MR 96e:14059 Zbl 0839.14008

[Carrell and Lieberman 1973] J. B. Carrell and D. I. Lieberman, "Holomorphic vector fields and Kaehler manifolds", *Invent. Math.* **21** (1973), 303–309. MR 48 #4356 Zbl 0253.32017

[Carrell and Lieberman 1977] J. B. Carrell and D. I. Lieberman, "Vector fields and Chern numbers", *Math. Ann.* **225**:3 (1977), 263–273. MR 55 #8416 Zbl 0365.32020

[Carrell and Sommese 1979] J. B. Carrell and A. J. Sommese, "Some topological aspects of $\mathbf{C}^*$ actions on compact Kaehler manifolds", *Comment. Math. Helv.* **54** (1979), 567–582. MR 80m:32032 Zbl 0466.32015

[Cartan 1951] H. Cartan, "La transgression dans un groupe de Lie et dans un espace fibré principal", pp. 57–71 in *Colloque de topologie (espaces fibrés), Bruxelles, 1950*, Georges Thone, Liège, 1951. MR 13,107f Zbl 0045.30701

[Feng 2003] H. Feng, "Holomorphic equivariant cohomology via a transversal holomorphic vector field", *Internat. J. Math.* **14**:5 (2003), 499–514. MR 2004j:32022 Zbl 1050.32013

[Frankel 1959] T. Frankel, "Fixed points and torsion on Kähler manifolds", *Ann. of Math.* (2) **70** (1959), 1–8. MR 24 #A1730 Zbl 0088.38002

[Fujiki 1979] A. Fujiki, "Fixed points of the actions on compact Kähler manifolds", *Publ. Res. Inst. Math. Sci.* **15**:3 (1979), 797–826. MR 81c:32055 Zbl 0446.53021

[Ginzburg 1987] V. A. Ginzburg, "Equivariant cohomology and Kähler geometry", *Funktsional. Anal. i Prilozhen.* **21**:4 (1987), 19–34. In Russian; translated in *Funct. Anal. Appl.* **21**:4 (1987), 271-283. MR 89b:58013 Zbl 0656.53062

[Goresky et al. 1998] M. Goresky, R. Kottwitz, and R. MacPherson, "Equivariant cohomology, Koszul duality, and the localization theorem", *Invent. Math.* **131**:1 (1998), 25–83. MR 99c:55009 Zbl 0897.22009

[Kaveh 2005] K. Kaveh, "Vector fields and the cohomology ring of toric varieties", *Canad. Math. Bull.* **48**:3 (2005), 414–427. MR 2007b:14111 Zbl 1094.14035

[Kirwan 1984] F. C. Kirwan, *Cohomology of quotients in symplectic and algebraic geometry*, Mathematical Notes **31**, Princeton University Press, Princeton, NJ, 1984. MR 86i:58050 Zbl 0553.14020

[Lillywhite 2003] S. Lillywhite, "Formality in an equivariant setting", *Trans. Amer. Math. Soc.* **355**:7 (2003), 2771–2793. MR 2004e:55016 Zbl 1021.55006

[Puppe 1974] V. Puppe, "On a conjecture of Bredon", *Manuscripta Math.* **12** (1974), 11–16. MR 49 #11502 Zbl 0277.57011

[Puppe 1979/80] V. Puppe, "Deformations of algebras and cohomology of fixed point sets", *Manuscripta Math.* **30**:2 (1979/80), 119–136. MR 81e:57038 Zbl 0418.57015

[Teleman 2000] C. Teleman, "The quantization conjecture revisited", *Ann. of Math.* (2) **152**:1 (2000), 1–43. MR 2002d:14073 Zbl 0980.53102

Received March 30, 2006. Revised October 20, 2006.

JIM CARRELL
DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF BRITISH COLUMBIA
ROOM 121, 1984 MATHEMATICS ROAD
VANCOUVER, B.C.
CANADA V6T 1Z2
carrell@math.ubc.ca

KIUMARS KAVEH
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
40 ST. GEORGE STREET
TORONTO, ONTARIO
CANADA M5S 2E4
kaveh@math.utoronto.ca

VOLKER PUPPE
UNIVERSITÄT KONSTANZ
FACHBEREICH MATHEMATIK UND STATISTIK
FACH D197
D-78457 KONSTANZ
GERMANY
volker.puppe@uni-konstanz.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

UNIQUENESS OF THE CHEEGER SET OF A CONVEX BODY

VICENT CASELLES, ANTONIN CHAMBOLLE AND MATTEO NOVAGA

We prove that if $C \subset \mathbb{R}^N$ is of class C^2 and uniformly convex, the Cheeger set of C is unique. The Cheeger set of C is the set that minimizes, inside C, the ratio of perimeter over volume.

1. Introduction

For a nonempty open bounded subset Ω of $\mathbb{R}^N$, the *Cheeger constant* of Ω is the quantity

$$h_\Omega = \min_{K \subseteq \Omega} \frac{P(K)}{|K|}. \tag{1}$$

Here $|K|$ denotes the N-dimensional volume of K and $P(K)$ denotes the perimeter of K. The minimum in (1) is taken over all nonempty sets of finite perimeter contained in Ω. A *Cheeger set* of Ω is any set $G \subseteq \Omega$ which minimizes (1). If Ω minimizes (1), we say that it is Cheeger in itself. We observe that the minimum in (1) is attained at a subset G of Ω such that ∂G intersects $\partial\Omega$: otherwise we could diminish the quotient $P(G)/|G|$ by dilating G.

For any set K of finite perimeter in $\mathbb{R}^N$, define

$$\lambda_K := \frac{P(K)}{|K|}.$$

Thus $\lambda_G = h_G$ for any Cheeger set G of Ω. Moreover, G is a Cheeger set of Ω if and only if G minimizes

$$\min_{K \subseteq \Omega} P(K) - \lambda_G |K|. \tag{2}$$

We say that a set $\Omega \subset \mathbb{R}^N$ is *calibrable* if Ω minimizes the problem

$$\min_{K \subseteq \Omega} P(K) - \lambda_\Omega |K|. \tag{3}$$

MSC2000: primary 35J70; secondary 49J40, 52A20.
Keywords: convex bodies, variational problems.
Caselles was partially support by the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya and by PNPGC project, reference BFM2003-02125.

Any Cheeger set G of Ω is clearly calibrable. Thus, Ω is a Cheeger set of itself if and only if it is calibrable.

Finding the Cheeger sets of a given Ω is a difficult task. The task is simplified if Ω is a convex set and $N = 2$. In that case, the Cheeger set of Ω is unique and equals the set $\Omega^R \oplus B(0, R)$, where $\Omega^R := \{x \in \Omega : \mathrm{dist}(x, \partial\Omega) > R\}$ is such that $|\Omega^R| = \pi R^2$ and $A \oplus B := \{a + b : a \in A,\ b \in B\}$, for $A, B \subset \mathbb{R}^2$ [Alter et al. 2005b; Kawohl and Lachand-Robert 2006]. In particular, in this case the Cheeger set is convex.

A convex set $\Omega \subseteq \mathbb{R}^2$ is Cheeger in itself if and only if $\operatorname{ess\,sup}_{x\in\partial\Omega}\kappa_\Omega(x) \le \lambda_\Omega$, where $\kappa_\Omega(x)$ denotes the curvature of $\partial\Omega$ at the point x. This has been proved in [Giusti 1978; Bellettini et al. 2002; Kawohl and Lachand-Robert 2006; Alter et al. 2005b; Kawohl and Novaga 2006], though it was stated in terms of calibrability in the second and fourth of these references. The proof in [Giusti 1978] had a complementary result: if Ω is Cheeger in itself then Ω is strictly calibrable, that is, for any set $K \subsetneq \Omega$, we have

$$0 = P(\Omega) - \lambda_\Omega|\Omega| < P(K) - \lambda_\Omega|K|.$$

(This implies that the gravity-less capillary problem with vertical contact angle at the boundary, given by

$$\begin{aligned} -\mathrm{div}\frac{Du}{\sqrt{1+|Du|^2}} &= \lambda_\Omega \quad \text{in } \Omega,\\ -\frac{Du}{\sqrt{1+|Du|^2}}\cdot \nu^\Omega &= 1 \qquad \text{in } \partial\Omega, \end{aligned} \tag{4}$$

has a solution. Indeed, the two problems are equivalent [Giusti 1978; Kawohl and Kutev 1995].)

Our purpose in this paper is to extend the preceding result to $\mathbb{R}^N$, that is, to prove the uniqueness and convexity of the Cheeger set contained in a convex set $\Omega \subset \mathbb{R}^N$. We have to assume, in addition, that Ω is uniformly convex and of class C^2. This regularity assumption is probably too strong, and its removal is the subject of current research [Alter and Caselles 2007]. The characterization of a convex set $\Omega \subset \mathbb{R}^N$ of class $C^{1,1}$ which is Cheeger in itself (also called calibrable) in terms of the mean curvature of its boundary was proved in [Alter et al. 2005a]. The precise result states that such a set Ω is Cheeger in itself if and only if $\kappa_\Omega(x) \le \lambda_\Omega$ for almost any $x \in \partial\Omega$, where $\kappa_\Omega(x)$ denotes the sum of the principal curvatures of the boundary of Ω, which is to say, $N-1$ times the mean curvature of $\partial\Omega$ at x. In [Alter et al. 2005a] it was also proved that for any convex set $\Omega \subset \mathbb{R}^N$ there exists a maximal Cheeger set contained in Ω which is convex. These results were extended to convex sets Ω satisfying a regularity condition and anisotropic norms in $\mathbb{R}^N$ (including the crystalline case) in [Caselles et al. 2005].

In particular, we obtain that $\Omega \subset \mathbb{R}^N$ is the unique Cheeger set of itself, whenever Ω is a C^2, uniformly convex calibrable set. We point out that, by Theorems 1.1 and 4.2 in [Giusti 1978], this uniqueness result is equivalent to the existence of a solution $u \in W^{1,\infty}_{\mathrm{loc}}(\Omega)$ of the capillary problem (4).

In Section 2 we collect some definitions and recall results about the mean curvature operator in (4) and the subdifferential of the total variation. In Section 3 we state and prove our uniqueness result.

2. Preliminaries

2.1. ***BV functions.*** Let Ω be an open subset of $\mathbb{R}^N$. A function $u \in L^1(\Omega)$ whose gradient Du in the sense of distributions is a (vector valued) Radon measure with finite total variation in Ω is called a function of bounded variation. The class of such functions will be denoted by $BV(\Omega)$. The total variation of Du on Ω turns out to be

$$\sup\left\{\int_\Omega u \operatorname{div} z\, dx : z \in C_0^\infty(\Omega;\mathbb{R}^N),\ \|z\|_{L^\infty(\Omega)} := \operatorname{ess\,sup}_{x\in\Omega}|z(x)| \le 1\right\}, \tag{5}$$

(where for a vector $v = (v_1, \dots, v_N) \in \mathbb{R}^N$ we set $|v|^2 := \sum_{i=1}^N v_i^2$) and will be denoted by $|Du|(\Omega)$ or by $\int_\Omega |Du|$. The map $u \mapsto |Du|(\Omega)$ is $L^1_{\mathrm{loc}}(\Omega)$-lower semicontinuous. $BV(\Omega)$ is a Banach space when endowed with the norm $\int_\Omega |u|\, dx + |Du|(\Omega)$. We recall that $BV(\mathbb{R}^N) \subseteq L^{N/(N-1)}(\mathbb{R}^N)$.

A measurable set $E \subseteq \mathbb{R}^N$ is said to be of finite perimeter in $\mathbb{R}^N$ if (5) is finite when we substitute for u the characteristic function χ_E of E and $\Omega = \mathbb{R}^N$. The perimeter of E is defined as $P(E) := |D\chi_E|(\mathbb{R}^N)$. For more information on functions of bounded variation we refer to [Ambrosio et al. 2000].

Finally, we denote by $\mathcal{H}^{N-1}$ the $(N-1)$-dimensional Hausdorff measure. We recall that when E is a finite-perimeter set with regular boundary (for instance, Lipschitz), its perimeter $P(E)$ also coincides with the more standard definition $\mathcal{H}^{N-1}(\partial E)$.

2.2. ***A generalized Green's formula.*** Let Ω be an open subset of $\mathbb{R}^N$. Following [Anzellotti 1983a], let

$$X_2(\Omega) := \{z \in L^\infty(\Omega;\mathbb{R}^N) : \operatorname{div} z \in L^2(\Omega)\}.$$

If $z \in X_2(\Omega)$ and $w \in L^2(\Omega) \cap BV(\Omega)$ we define the functional

$$(z \cdot Dw) : C_0^\infty(\Omega) \to \mathbb{R}$$

by the formula

$$\langle (z \cdot Dw), \varphi\rangle := -\int_\Omega w\,\varphi \operatorname{div} z\, dx - \int_\Omega w\, z \cdot \nabla\varphi\, dx.$$

Then $(z \cdot Dw)$ is a Radon measure in Ω,

$$\int_\Omega (z \cdot Dw) = \int_\Omega z \cdot \nabla w \, dx \quad \text{for } w \in L^2(\Omega) \cap W^{1,1}(\Omega).$$

Recall that the outer unit normal to a point $x \in \partial\Omega$ is denoted by $\nu^\Omega(x)$. We recall the following result proved in [Anzellotti 1983a].

Theorem 1. *Let $\Omega \subset \mathbb{R}^N$ be a bounded open set with Lipschitz boundary. Let $z \in X_2(\Omega)$. Then there exists a function $[z \cdot \nu^\Omega] \in L^\infty(\partial\Omega)$ satisfying $\|[z \cdot \nu^\Omega]\|_{L^\infty(\partial\Omega)} \leq \|z\|_{L^\infty(\Omega;\mathbb{R}^N)}$, and such that for any $u \in BV(\Omega) \cap L^2(\Omega)$ we have*

$$\int_\Omega u \operatorname{div} z \, dx + \int_\Omega (z \cdot Du) = \int_{\partial\Omega} [z \cdot \nu^\Omega] u \, d\mathcal{H}^{N-1}.$$

Moreover, if $\varphi \in C^1(\overline{\Omega})$ then $[(\varphi z) \cdot \nu^\Omega] = \varphi[z \cdot \nu^\Omega]$.

This result is complemented with the following.

Theorem 2 [Anzellotti 1983b]. *Let $\Omega \subset \mathbb{R}^N$ be a bounded open set with a boundary of class C^1. Let $z \in C(\overline{\Omega}; \mathbb{R}^N)$ with* div $z \in L^2(\Omega)$. *Then*

$$[z \cdot \nu^\Omega](x) = z(x) \cdot \nu^\Omega(x) \qquad \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega.$$

2.3. ***Some auxiliary results.*** Let Ω be an open bounded subset of $\mathbb{R}^N$ with Lipschitz boundary, and let $\varphi \in L^1(\Omega)$. For all $\epsilon > 0$, we let $\Psi_\varphi^\varepsilon : L^2(\Omega) \to (-\infty, +\infty]$ be the functional defined by

$$\Psi_\varphi^\epsilon(u) := \begin{cases} \displaystyle\int_\Omega \sqrt{\epsilon^2 + |Du|^2} + \int_{\partial\Omega} |u - \varphi| & \text{if } u \in L^2(\Omega) \cap BV(\Omega), \\ +\infty & \text{if } u \in L^2(\Omega) \setminus BV(\Omega). \end{cases} \tag{6}$$

As it is proved in [Giusti 1976], if $f \in W^{1,\infty}(\Omega)$, then the minimum $u \in BV(\Omega)$ of the functional

$$\Psi_\varphi^\epsilon(u) + \int_\Omega |u(x) - f(x)|^2 \, dx \tag{7}$$

belongs to $u \in C^{2+\alpha}(\Omega)$, for every $\alpha < 1$. The minimum u of (7) is a solution of

$$\begin{cases} u - \dfrac{1}{\lambda} \operatorname{div} \dfrac{Du}{\sqrt{\varepsilon^2 + |Du|^2}} = f(x) & \text{in } \Omega, \\ u = \varphi & \text{on } \partial\Omega, \end{cases} \tag{8}$$

where the boundary condition is taken in a generalized sense [Lichnewsky and Temam 1978], i.e.,

$$\left[\frac{Du}{\sqrt{\varepsilon^2 + |Du|^2}} \cdot \nu^\Omega \right] \in \operatorname{sign}(\varphi - u) \qquad \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega.$$

Observe that (8) can be written as

$$u + \frac{1}{\lambda}\partial\Psi^{\epsilon}_{\varphi}(u) \ni f. \tag{9}$$

We are particularly interested in the case where $\varphi = 0$. As we shall show below (see also [Alter et al. 2005a]) in the case of interest to us we have $u > 0$ on $\partial\Omega$ and thus,

$$\left[\frac{Du}{\sqrt{\varepsilon^2 + |Du|^2}} \cdot \nu^{\Omega}\right] = -1 \qquad \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega.$$

It follows that u is a solution of the first equation in (8) with vertical contact angle at the boundary.

As $\epsilon \to 0^+$, the solution of (8) converges to the solution of

$$\begin{cases} u + \dfrac{1}{\lambda}\partial\Psi_{\varphi}(u) = f(x) & \text{in } \Omega, \\ u = \varphi & \text{on } \partial\Omega. \end{cases} \tag{10}$$

where $\Psi : L^2(\Omega) \to (-\infty, +\infty]$ is given by

$$\Psi_{\varphi}(u) := \begin{cases} \displaystyle\int_{\mathbb{R}^N} |Du| + \int_{\partial\Omega} |u - \varphi| & \text{if } u \in L^2(\Omega) \cap BV(\Omega), \\ +\infty & \text{if } u \in L^2(\Omega) \setminus BV(\Omega). \end{cases} \tag{11}$$

In this case $\partial\Psi_{\varphi}$ represents the operator $-\operatorname{div} \dfrac{Du}{|Du|}$ with the boundary condition $u = \varphi$ in $\partial\Omega$, as shown by:

Lemma 2.1 [Andreu et al. 2001]. *The following assertions are equivalent*:

(a) $v \in \partial\Psi_{\varphi}(u)$.

(b) $u \in L^2(\Omega) \cap BV(\Omega)$, $v \in L^2(\Omega)$, *and there exists* $z \in X_2(\Omega)$ *with* $\|z\|_{\infty} \le 1$, *such that* $v = -\operatorname{div} z$ *in* $\mathcal{D}'(\Omega)$, $z \cdot Du = |Du|$, *and*

$$[z \cdot \nu^{\Omega}] \in \operatorname{sign}(\varphi - u) \qquad \mathcal{H}^{N-1}\text{-a.e. on } \partial\Omega.$$

Notice that the solution $u \in L^2(\Omega)$ of (10) minimizes the problem

$$\min_{u \in BV(\Omega)} \int_{\Omega} |Du| + \int_{\partial\Omega} |u(x) - \varphi(x)|\, d\mathcal{H}^{N-1}(x) + \frac{\lambda}{2}\int_{\Omega} |u(x) - f(x)|^2\, dx, \tag{12}$$

and the two problems are equivalent.

3. The uniqueness theorem

We now state our main result.

Theorem 3. *Let C be a convex body in $\mathbb{R}^N$. Assume that C is uniformly convex, with boundary of class C^2. Then the Cheeger set of C is convex and unique.*

We do not believe that the regularity and the uniform convexity of C is essential for this result (see [Alter and Caselles 2007]).

Theorem 4 [Alter et al. 2005a, Theorems 6 and 8, Proposition 4]. *Let C be a convex body in $\mathbb{R}^N$ with boundary of class $C^{1,1}$. For any $\lambda, \varepsilon > 0$, there is a unique solution u_ε of the equation*

$$\begin{cases} u_\varepsilon - \frac{1}{\lambda}\operatorname{div} \dfrac{Du_\varepsilon}{\sqrt{\varepsilon^2 + |Du_\varepsilon|^2}} = 1 & \text{in } C, \\ u_\varepsilon = 0 & \text{on } \partial C \end{cases} \tag{13}$$

such that $0 \le u_\varepsilon \le 1$. Moreover, there exist λ_0 and ε_0, depending only on ∂C, such that if $\lambda \ge \lambda_0$ and $\varepsilon \le \varepsilon_0$, then u_ε is a concave function such that $u_\varepsilon \ge \alpha > 0$ on ∂C for some $\alpha > 0$. Hence, u_ε satisfies

$$\left[\frac{Du^\epsilon}{\sqrt{\epsilon^2 + |Du^\epsilon|^2}} \cdot \nu^C\right] = \operatorname{sign}(0 - u^\epsilon) = -1 \quad \text{on } \partial C. \tag{14}$$

As $\varepsilon \to 0$, the functions u_ε converge to the concave function u minimizing the problem

$$\min_{u \in BV(C)} \int_C |Du| + \int_{\partial C} |u(x)|\, d\mathcal{H}^{N-1}(x) + \frac{\lambda}{2}\int_C |u(x) - 1|^2\, dx; \tag{15}$$

equivalently, if u is extended with zero out of C, the extension minimizes

$$\int_{\mathbb{R}^N} |Du| + \frac{\lambda}{2}\int_{\mathbb{R}^N} |u - \chi_C|^2\, dx.$$

The function u satisfies $0 \le u < 1$. The superlevel set $\{u \ge t\}$, for $t \in (0, 1]$, is contained in C and minimizes the problem

$$\min_{F \subset C} P(F) - \lambda(1-t)|F|. \tag{16}$$

It was proved in [Alter et al. 2005a] (see also [Caselles et al. 2005]) that the set $C^* = \{u = \max_C u\}$ is the maximal Cheeger set contained in C, that is, the maximal set that solves (1). Moreover, one has $u = 1 - h_C/\lambda > 0$ in C^* and $h_C = \lambda_{C^*}$.

If we want to consider what happens inside C^*, and in particular whether there are other Cheeger sets, we have to analyze the level sets of u_ε before passing to the limit as $\epsilon \to 0^+$. To do this, we introduce the following rescaling of u_ε:

$$v_\varepsilon = \frac{u_\varepsilon - m_\varepsilon}{\varepsilon} \le 0,$$

where $m_\varepsilon = \max_C u_\varepsilon \to 1 - h_C/\lambda$ as $\varepsilon \to 0$. The function v_ε is a generalized solution of the equation:

$$\begin{cases} \varepsilon v_\varepsilon - \dfrac{1}{\lambda}\operatorname{div} \dfrac{Dv_\varepsilon}{\sqrt{1+|Dv_\varepsilon|^2}} = 1 - m_\varepsilon & \text{in } C, \\ v_\varepsilon = -\dfrac{m_\varepsilon}{\varepsilon} & \text{on } \partial C. \end{cases} \tag{17}$$

We define the vector field

$$z_\varepsilon = Du_\varepsilon/\sqrt{\varepsilon^2 + |Du_\varepsilon|^2} = Dv_\varepsilon/\sqrt{1+|Dv_\varepsilon|^2};$$

it lies in $L^\infty(C)$, has uniformly bounded divergence, and satisfies $|z_\varepsilon| \le 1$ a.e. in C and, by (14), $[z_\varepsilon \cdot \nu_C] = -1$ on ∂C.

We now study the limit of v_ε and z_ε as $\varepsilon \to 0$. By the concavity of v_ε, for each $\varepsilon > 0$ small enough and each $s \in (0, |C|)$, there exists a (convex) superlevel set C_s^ε of v_ε such that $|C_s^\varepsilon| = s$. Moreover, $\{v_\varepsilon = 0\}$ is a null set: otherwise, since v_ε is concave, it would be a convex set of positive measure, hence with nonempty interior. We would then have $v_\varepsilon = \operatorname{div} z_\varepsilon = 0$, hence $1 - m_\varepsilon = 0$ in the interior of $\{v_\varepsilon = 0\}$. This is a contradiction with Theorem 4 for $\varepsilon > 0$ small enough.

Hence we may take $C_0^\varepsilon := \{v_\varepsilon = 0\}$ and $C_{|C|}^\varepsilon := C$. The boundaries $\partial C_s^\varepsilon \cap C$ define a foliation in C, in the sense that for all $x \in C$, there exists a unique value of $s \in [0, |C|]$ such that $x \in \partial C_s^\varepsilon$.

A sequence of uniformly bounded convex sets is compact both for the L^1 and Hausdorff topologies. Hence, up to a subsequence, we may assume that the C_s^ε converge to convex sets C_s, each of volume s, first for any $s \in \mathbb{Q} \cap (0, |C|)$ and then by continuity for any s. Possibly extracting a further subsequence, we may assume that there exists $s_* \in [0, |C|]$ such that v_ε goes to a concave function v in C_s for any $s < s_*$, and to $-\infty$ outside $C_* := C_{s_*}$. We may also assume that $z_\varepsilon \rightharpoonup z$ weakly* in $L^\infty(C)$, for some vector field z satisfying $|z| \le 1$ a.e. in C. From (13) we have in the limit

$$-\operatorname{div} z = \lambda(1-u) \qquad \text{in } \mathcal{D}'(C). \tag{18}$$

Moreover, $-\operatorname{div} z \in \partial\Psi_0(u)$ by the results recalled in Section 2. We then see from (18) that

$$-\operatorname{div} z = h_C \qquad \text{in } C^*, \tag{19}$$

while $-\operatorname{div} z > h_C$ a.e. on $C \setminus C^*$.

Set $s^* := |C^*|$, so $C^* = C_{s^*}$. By Theorem 4, for $s \ge s^*$, the set C_s is a minimizer of the variational problem

$$\min_{E \subseteq C} P(E) - \mu_s |E|, \tag{20}$$

for some $\mu_s \geq h_C$ (μ_s is equal to the constant value of $-\operatorname{div} z = \lambda(1-u)$ on $\partial C_s \cap C$; see (16)). Notice that μ_s is bounded from above by $P(C)/(|C|-s)$: indeed,

$$-\int_{C\setminus C_s^\varepsilon} \operatorname{div} z_\varepsilon(x)\,dx = \mathcal{H}^{N-1}(\partial C \setminus \partial C_s^\varepsilon) - \int_{\partial C_s^\varepsilon \cap C} \frac{|Du_\varepsilon|}{\sqrt{1+|Du_\varepsilon|^2}} \;\leq\; P(C)$$

for $\varepsilon > 0$, since the inner normal to C_s^ε at $x \in \partial C_s^\varepsilon \cap C$ is $Du_\varepsilon(x)/|Du_\varepsilon(x)|$. On the other hand,

$$-\int_{C\setminus C_s^\varepsilon} \operatorname{div} z_\varepsilon(x)\,dx = \int_{C\setminus C_s^\varepsilon} \lambda(1-u_\varepsilon(x))\,dx \geq \mu_s^\varepsilon(|C|-s),$$

where μ_s^ε is the constant value of $\lambda(1-u_\varepsilon)$ on the level set $\partial C_s^\varepsilon \cap C$, and goes to μ_s as $\varepsilon \to 0$. A more careful analysis would show, in fact, that

$$\mu_s \leq \frac{P(C)-P(C_s)}{|C|-s}.$$

For $s > s^*$, we have $\mu_s > h_C$ and the set C_s is the unique minimizer of the variational problem (20). As a consequence (see [Alter et al. 2005a; Caselles et al. 2005]) for any $s > s^*$ the set C_s is also the unique minimizer of $P(E)$ among all $E \subseteq C$ of volume s.

Lemma 3.1. *We have $s_* > 0$ and the sets C_s are Cheeger sets in C for any $s \in [s_*, s^*]$.*

Proof. Let $s_* < s \leq |C|$. If $x \in \partial C_s^\epsilon \setminus \partial C$, then

$$0 - v_\varepsilon(x) \;\leq\; Dv_\varepsilon(x)\cdot(\bar{x}_\varepsilon - x)$$

where $v_\varepsilon(\bar{x}_\varepsilon) = \max_C v_\varepsilon$. Hence, $\lim_{\varepsilon\to 0}\inf_{\partial C_s^\varepsilon \setminus \partial C} |Dv_\varepsilon| = +\infty$. Since $[z_\varepsilon \cdot \nu^C] = -1$ on ∂C and $P(C_s^\varepsilon) \to P(C_s)$, we deduce

$$\begin{aligned}
-\int_{\partial C_s^\varepsilon} & [z_\varepsilon(x)\cdot \nu^{C_s^\varepsilon}(x)]\,d\mathcal{H}^{N-1}(x) \\
&= \int_{\partial C_s^\varepsilon \setminus \partial C} \frac{|Dv_\varepsilon(x)|}{\sqrt{1+|Dv_\varepsilon(x)|^2}}\,d\mathcal{H}^{N-1}(x) \;+\; \mathcal{H}^{N-1}(\partial C_s^\varepsilon \cap \partial C) \;\to\; P(C_s)
\end{aligned}$$

as $\varepsilon \to 0^+$. Hence,

$$\begin{aligned}
\int_{\partial C_s} \left[z\cdot \nu^{C_s}\right] d\mathcal{H}^{N-1} &= \int_{C_s} \operatorname{div} z = \lim_{\varepsilon\to 0} \int_{C_s^\varepsilon} \operatorname{div} z_\varepsilon \\
&= \lim_{\varepsilon\to 0} \int_{\partial C_s^\varepsilon} [z_\varepsilon \cdot \nu_{C_s^\varepsilon}]\,d\mathcal{H}^{N-1} = -\,P(C_s).
\end{aligned}$$

Since $|z| \le 1$ a.e. in C, we deduce that $[z \cdot \nu^{C_s}] = -1$ on ∂C_s for any $s > s_*$ (in particular, $|z| = 1$ a.e. in $C \setminus C_*$). Using this and (19), we have for all $s_* < s \le s^*$

$$\frac{P(C_s)}{|C_s|} = h_C. \tag{21}$$

This has two consequences. First, from the isoperimetric inequality, we obtain

$$h_C = \frac{P(C_s)}{|C_s|} \ge \frac{P(B_1)}{\big(|B_1|^{N-1} s\big)^{1/N}},$$

if $s \in (s_*, s^*]$, so that $s_* > 0$. Moreover, C_s is a Cheeger set for any $s \in (s_*, s^*]$, and by continuity C_* is also a Cheeger set. □

Since the sets C_s are convex minimizers of $P(E) - \mu_s|E|$ among all $E \subseteq C$, for $s \ge s_*$, their boundary is of class $C^{1,1}$ [Brézis and Kinderlehrer 1974; Stredulinsky and Ziemer 1997], with curvature at most μ_s, and equal to μ_s in the interior of C (note that $\mu_s = h_C$ for $s \in [s_*, s^*]$).

Remark 3.2. Either $s_* = s^*$, and so $C_* = C^*$, or $s_* < s^*$, and so $C^* = \bigcup_{s \in (s_*, s^*)} C_s$. In the latter case, the supremum of the sum κ_{C^*} of the principal curvatures on ∂C^* is equal to h_C. Indeed, if this were not the case, by considering $C' \subset \operatorname{int}(C^*)$ with curvature strictly below h_C, together with the smallest set C_s with $s > s_*$ containing C', we would get $\kappa_{C'}(x) \ge \kappa_{C_s}(x) = h_C$ at all $x \in \partial C' \cap \partial C_s$, a contradiction. In particular, $C = C_*$ if the supremum of κ_C on ∂C is strictly less than $P(C)/|C|$; this condition also implies $C = C^*$ by [Alter et al. 2005a].

From the strong convergence of Dv_ε to Dv (in $L^2(C_s)$ for any $s < s_*$), we deduce that $z = Dv/\sqrt{1 + |Dv|^2}$ in C_*. It follows that v satisfies the equation

$$-\operatorname{div} \frac{Dv}{\sqrt{1 + |Dv|^2}} = h_C \quad \text{in } C_*. \tag{22}$$

Integrating both sides of (22) in C_*, we deduce that

$$\left[\frac{Dv}{\sqrt{1 + |Dv|^2}} \cdot \nu^{C_*} \right] = -1 \quad \text{on } \partial C_*.$$

Lemma 3.3. *The set C_* is the **minimal** Cheeger set of C; that is, any Cheeger set of C must contain C_*.*

Proof. Let $K \subseteq C^*$ be a Cheeger set in C. We have

$$h_C|K| = -\int_K \operatorname{div} z = -\int_{\partial K} [z \cdot \nu^K] \, d\mathcal{H}^{N-1} = P(K),$$

so $[z \cdot \nu^K] = -1$ a.e. on ∂K. Let ν^ϵ and ν be the vector fields of unit normals to the sets C_s^ϵ and C_s, $s \in [0, |C|]$, respectively. By the Hausdorff convergence of C_s^ϵ

to C_s as $\epsilon \to 0^+$ for any $s \in [0, |C|]$, we have $\nu^\epsilon \to \nu$ a.e. in C. On the other hand, $|z_\epsilon + \nu^\epsilon| \to 0$ locally uniformly in $C \setminus \bar{C}_*$: indeed, in C,

$$|z_\epsilon + \nu^\epsilon| = \left| \frac{Dv_\varepsilon}{\sqrt{1+|Dv_\varepsilon|^2}} - \frac{Dv_\varepsilon}{|Dv_\varepsilon|} \right| = \left| \frac{|Dv_\varepsilon|}{\sqrt{1+|Dv_\varepsilon|^2}} - 1 \right|.$$

Since $|Dv_\epsilon| \to \infty$ uniformly in any subset of C at positive distance from C_* (see the first lines of the proof of Lemma 3.1), this shows the uniform convergence of $|z_\epsilon + \nu^\epsilon|$ to 0 in such subsets.

These two facts imply that $z = -\nu$ a.e. on $C \setminus C_*$. By modifying z in a set of null measure, we may assume that $z = -\nu$ on $C \setminus C_*$. We recall that the sets C_s, $s \geq s_*$ are minimizers of variational problems of the form $\min_{K \subseteq C} P(K) - \mu|K|$, for some values of μ (with $\mu = h_C$ as long as $s \leq s^*$ and $\mu = \mu_s > h_C$ continuously increasing with $s > s^*$). Since these sets are convex, with boundary (locally) uniformly of class $C^{1,1}$, and the map $s \to C_s$ is continuous in the Hausdorff topology, we conclude that the normal $\nu(x)$ is a continuous function in $C \setminus \mathrm{int}(C_*)$.

Since $|z| < 1$ inside C_* and $[z \cdot \nu^K] = -1$ a.e. on ∂K, by [Anzellotti 1983a, Theorem 1]) we have that the boundary of K must be outside the interior of C_*, hence either $K \supseteq C_*$ or $K \cap C_* = \varnothing$ (modulo a null set). Let us prove that the last situation is impossible. Indeed, assume that $K \cap C_* = \varnothing$ (modulo a null set). Since ∂K is of class C^1 out of a closed set of zero $\mathcal{H}^{N-1}$-measure (see [Gonzalez et al. 1983]) and z is continuous in $C \setminus \mathrm{int}(C_*)$, by Theorem 2 we have

$$z(x) \cdot \nu^K(x) = -1 \qquad \mathcal{H}^{N-1}\text{-a.e. on } \partial K. \tag{23}$$

Now, since $K \cap C_* = \varnothing$ (modulo a null set), then there is some $s \geq s_*$ and some $x \in \partial C_s \cap \partial K$ such that $\nu^K(x) + \nu(x) = 0$. Fix $0 < \epsilon < 2$. By a slight perturbation, if necessary, we may assume that $x \in \partial C_s \cap \partial K$ with $s > s_*$, (23) holds at x and

$$|\nu^K(x) + \nu(x)| < \epsilon. \tag{24}$$

Since by (23) we have $\nu(x) = -z(x) = \nu^K(x)$ we obtain a contradiction with (24). We deduce that $K \supseteq C_*$. □

Therefore, in order to prove the uniqueness of the Cheeger set of C, it is enough to show that

$$C_* = C^*. \tag{25}$$

Recall that the boundary of both C_* and C^* is of class $C^{1,1}$, and the sum of its principal curvatures is less than or equal h_C, and constantly equal to h_C in the interior of C. We now show that if $C_* \neq C^*$ and under additional assumptions, the sum of the principal curvatures of the boundary of C^* (or of any C_s for $s \in (s_*, s^*]$) must be h_C out of C_*.

Lemma 3.4. *Assume that C has C^2 boundary. Let $s \in (s_*, s^*]$ and $x \in \partial C_s \setminus \partial C_*$. If the sum of the principal curvatures of ∂C_s at x is strictly below h_C, then the Gaussian curvature of ∂C at x is* 0*.*

Proof. Let $x \in \partial C_s \setminus \partial C_*$ and assume the sum of the principal curvatures of ∂C_s at x is strictly below h_C (assuming x is a Lebesgue point for the curvature on ∂C_s). Necessarily, this implies that $x \in \partial C$. Assume then that the Gauss curvature of ∂C at x is positive: by continuity, in a neighborhood of x, C is uniformly convex and the sum of the principal curvatures is less than h_C. We may assume that near x, ∂C is the graph of a nonnegative, C^2 and convex function $f : B \to \mathbb{R}$ where B is an $(N-1)$-dimensional ball centered at x. We may as well assume that ∂C_s is the graph of $f_s : B \to \mathbb{R}$, which is $C^{1,1}$ [Brézis and Kinderlehrer 1974; Stredulinsky and Ziemer 1997], and also nonnegative and convex. In B, we have $f_s \geq f \geq 0$, and

$$D^2 f \;\geq\; \alpha I \quad \text{and} \quad \operatorname{div} \frac{Df}{\sqrt{1+|Df|^2}} = h$$

with $h \in C^0(\bar{B})$, $h < h_C$, $\alpha > 0$, while

$$\operatorname{div} \frac{Df_s}{\sqrt{1+|Df_s|^2}} = h\chi_{\{f=f_s\}} + h_C\chi_{\{f_s>f\}}$$

(where $\chi_{\{f=f_s\}}$ has positive density at x).

We let $g = f_s - f \geq 0$. Introducing the Lagrangian $\Psi : \mathbb{R}^{N-1} \to [0, +\infty)$ given by $\Psi(p) = \sqrt{1+|p|^2}$, we obtain, for a.e. $y \in B$,

$$\begin{aligned}
&(h_C - h(y))\chi_{\{g>0\}}(y) \\
&\qquad = \operatorname{div}\big(D\Psi(Df_s(y)) - D\Psi(Df(y))\big) \\
&\qquad = \operatorname{div}\left(\left(\int_0^1 D^2\Psi(Df(y) + t(Df_s(y) - Df(y)))\,dt\right) Dg(y)\right),
\end{aligned}$$

so that, letting $A(y) := \int_0^1 D^2\Psi(Df(y) + t\,Dg(y))\,dt$ (which is a positive definite matrix and Lipschitz continuous inside B), we see that g is the minimizer of the functional

$$w \mapsto \int_B \big(A(y)Dw(y)\cdot Dw(y) + (h_C - h(y))w(y)\big)\,dy$$

under the constraint $w \geq 0$ and with boundary condition $w = f_s - f$ on ∂B. Adapting the results in [Caffarelli and Rivière 1976] we get that $\{f = f_s\} = \{g = 0\}$ is the closure of a nonempty open set with boundary of zero $\mathcal{H}^{N-1}$-measure.

We therefore have found an open subset $D \subset \partial C \cap \partial C_s$, disjoint from ∂C_*, on which C is uniformly convex, with curvature less than h_C. Let φ be a smooth, nonnegative function with compact support in D. One easily shows that if $\varepsilon > 0$ is small enough, $\partial C_s - \varepsilon\varphi\nu^{C_s}$ is the boundary of a set C'_ϵ which is still convex, with

$P(C'_\epsilon)/|C'_\epsilon| > P(C_s)/|C_s| = h_C$ (just differentiate the map $\epsilon \to P(C'_\epsilon)/|C'_\epsilon|$), and the sum of its principal curvatures is less than h_C. This implies that for $\epsilon > 0$ small enough, the set $C' := C'_\epsilon$ is calibrable [Alter et al. 2005a], which in turn implies that $\min_{K \subset C'} P(K)/|K| = P(C')/|C'|$. But this contradicts $C_* \subset C'$, which is true for ε small enough. □

Proof of Theorem 3. Assume that C is C^2 and uniformly convex. Let us prove that its Cheeger set is unique. Assume by contradiction that $C^* \neq C_*$. From Lemma 3.4 we have that the sum of the principal curvatures of ∂C^* is h_C outside of C_*.

Let now $\bar{x} \in \partial C^* \cap \partial C_*$ be such that $\partial C^* \cap B_\rho(\bar{x}) \neq \partial C_* \cap B_\rho(\bar{x})$ for all $\rho > 0$ ($\partial C^* \cap \partial C_* \neq \varnothing$ since otherwise both C^* and C_* would be balls, which is impossible). Letting T be the tangent hyperplane to ∂C^* at $\bar{x}$, we can write ∂C^* and ∂C_* as the graph of two positive convex functions v^* and v_*, respectively, over $T \cap B_\rho(\bar{x})$ for $\rho > 0$ small enough. Identifying $T \cap B_\rho(\bar{x})$ with $B_\rho \subset \mathbb{R}^{N-1}$, we have that $v_*, v^* : B_\rho \to \mathbb{R}$ both solve the equation

$$- \operatorname{div} \frac{Dv}{\sqrt{1+|Dv|^2}} = f, \tag{26}$$

for some function $f \in L^\infty(B_\rho)$. Moreover, it holds $v_* \geq v^*$, $v_*(0) = v^*(0)$ and $v_*(y) > v^*(y)$ for some $y \in B_\rho$. Notice that $f = \lambda_C$ in the (open) set where $v_* > v^*$, in particular both functions are smooth in this set. Let D be an open ball such that $\overline{D} \subset B_\rho$, $v_* > v^*$ on D and $v_*(y) = v^*(y)$ for some $y \in \partial D$. Notice that, since both v^* and v_* belong to $C^\infty(D) \cap C^1(\overline{D})$, the fact that $v_*(y) = v^*(y)$ also implies that $Dv_*(y) = Dv^*(y)$. In D, both functions solve (26) with $f = \lambda_C$. Letting $w = v_* - v^*$, we obtain $w(y) = 0$ and $Dw(y) = 0$, while $w > 0$ inside D. Recalling the function $\Psi(p) = \sqrt{1+|p|^2}$, we have, for any $x \in D$,

$$\begin{aligned} 0 &= \operatorname{div} \big(D\Psi(Dv_*(x)) - D\Psi(Dv^*(x)) \big) \\ &= \operatorname{div} \left(\left(\int_0^1 D^2\Psi(Dv^*(x) + t(Dv_*(x) - Dv^*(x)))\, dt \right) Dw(x) \right), \end{aligned}$$

so that w solves a linear, uniformly elliptic equation with smooth coefficients. Then Hopf's lemma [Gilbarg and Trudinger 1983] implies that $Dw(y) \cdot \nu_D(y) < 0$, a contradiction. Hence $C_* = C^*$. □

Remark 3.5. As a consequence of Theorem 3 and the results of [Giusti 1978], if C is of class C^2 and uniformly convex, Equation (22) has a solution on the whole of C, if and only if C is a Cheeger set of itself, i.e., if and only if the sum of the principal curvatures of ∂C is less than or equal to $P(C)/|C|$.

Remark 3.6. The results of this paper can be easily extended to the anisotropic setting (see [Caselles et al. 2005]) provided the anisotropy is smooth and uniformly elliptic.

References

[Alter and Caselles 2007] F. Alter and V. Caselles, "Uniqueness of the Cheeger set of a convex body", preprint, 2007, Available at benasque.ecm.ub.es/2007pde/talks_contr/072cheeger8.pdf.

[Alter et al. 2005a] F. Alter, V. Caselles, and A. Chambolle, "A characterization of convex calibrable sets in $\mathbb{R}^N$", *Math. Ann.* **332**:2 (2005), 329–366. MR 2006g:35091 Zbl 02172739

[Alter et al. 2005b] F. Alter, V. Caselles, and A. Chambolle, "Evolution of characteristic functions of convex sets in the plane by the minimizing total variation flow", *Interfaces Free Bound.* **7**:1 (2005), 29–53. MR 2006b:35154 Zbl 1084.49003

[Ambrosio et al. 2000] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, New York, 2000. MR 2003a:49002 Zbl 0957.49001

[Andreu et al. 2001] F. Andreu, C. Ballester, V. Caselles, and J. Mazón, "The Dirichlet problem for the total variation flow", *J. Funct. Anal.* **180**:2 (2001), 347–403. MR 2001k:35143 Zbl 0973.35109

[Anzellotti 1983a] G. Anzellotti, "Pairings between measures and bounded functions and compensated compactness", *Ann. Mat. Pura Appl.* (4) **135** (1983), 293–318 (1984). MR 85m:46042 Zbl 0572.46023

[Anzellotti 1983b] G. Anzellotti, "Traces of bounded vector fields and the divergence theorem", preprint, Dipartimento di Matematica, Università di Trento, 1983.

[Bellettini et al. 2002] G. Bellettini, V. Caselles, and M. Novaga, "The total variation flow in $\mathbb{R}^N$", *J. Differential Equations* **184**:2 (2002), 475–525. MR 2003g:35105 Zbl 1036.35099

[Brézis and Kinderlehrer 1974] H. Brézis and D. Kinderlehrer, "The smoothness of solutions to nonlinear variational inequalities", *Indiana Univ. Math. J.* **23** (1974), 831–844. MR 50 #13881 Zbl 0278.49011

[Caffarelli and Rivière 1976] L. A. Caffarelli and N. M. Rivière, "On the rectifiability of domains with finite perimeter", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3**:2 (1976), 177–186. MR 53 #14288 Zbl 0362.49031

[Caselles et al. 2005] V. Caselles, A. Chambolle, S. Moll, and M. Novaga, "A characterization of convex calibrable sets in $\mathbb{R}^N$ with respect to anisotropic norms", preprint, 2005.

[Gilbarg and Trudinger 1983] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Grundlehren der Math. Wissenschaften **224**, 1983. Corrected reprint, 1998. MR 2001k:35004 Zbl 0562.35001

[Giusti 1976] E. Giusti, "Boundary value problems for non-parametric surfaces of prescribed mean curvature", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3**:3 (1976), 501–548. MR 58 #2572 Zbl 0344.35036

[Giusti 1978] E. Giusti, "On the equation of surfaces of prescribed mean curvature. Existence and uniqueness without boundary conditions", *Invent. Math.* **46**:2 (1978), 111–137. MR 58 #7337 Zbl 0381.35035

[Gonzalez et al. 1983] E. Gonzalez, U. Massari, and I. Tamanini, "On the regularity of boundaries of sets minimizing perimeter with a volume constraint", *Indiana Univ. Math. J.* **32**:1 (1983), 25–37. MR 84d:49043 Zbl 0486.49024

[Kawohl and Kutev 1995] B. Kawohl and N. Kutev, "Global behaviour of solutions to a parabolic mean curvature equation", *Differential Integral Equations* **8**:8 (1995), 1923–1946. MR 96f:35088 Zbl 0844.35050

[Kawohl and Lachand-Robert 2006] B. Kawohl and T. Lachand-Robert, "Characterization of Cheeger sets for convex subsets of the plane", *Pacific J. Math.* **225**:1 (2006), 103–118. MR 2007e:52002 Zbl 05170768

[Kawohl and Novaga 2006] B. Kawohl and M. Novaga, “The p-Laplace eigenvalue problem as $p \to 1$ and Cheeger sets in a Finsler metric”, preprint, 2006, Available at http://cvgmt.sns.it/papers/kawnov06/KawNovFeb27.pdf.

[Lichnewsky and Temam 1978] A. Lichnewsky and R. Temam, “Pseudosolutions of the time-dependent minimal surface problem”, *J. Differential Equations* **30**:3 (1978), 340–364. MR 80a:35059 Zbl 0368.49016

[Stredulinsky and Ziemer 1997] E. Stredulinsky and W. P. Ziemer, “Area minimizing sets subject to a volume constraint in a convex set”, *J. Geom. Anal.* **7**:4 (1997), 653–677. MR 99k:49089 Zbl 0940.49025

Received April 6, 2006. Revised July 21, 2006.

VICENT CASELLES
DEPARTAMENT DE TECNOLOGIA
UNIVERSITAT POMPEU FABRA
PASSEIG CIRCUMVALLACIÓ, 8
08003 BARCELONA
SPAIN
vicent.caselles@upf.edu

ANTONIN CHAMBOLLE
CMAP, ÉCOLE POLYTECHNIQUE
CNRS
91128 PALAISEAU
FRANCE
antonin.chambolle@polytechnique.fr

MATTEO NOVAGA
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI PISA
LARGO BRUNO PONTECORVO 5
56127 PISA
ITALY
novaga@dm.unipi.it

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

DIMENSION ESTIMATE OF HARMONIC FORMS ON COMPLETE MANIFOLDS

JUI-TANG RAY CHEN AND CHIUNG-JUE ANNA SUNG

We consider the space of polynomial-growth harmonic forms. We prove that the dimension of such spaces must be finite and can be estimated if the metric is uniformly equivalent to one with asymptotically nonnegative curvature operator. This implies that the space of harmonic forms of polynomial growth order on the connected sum manifolds with nonnegative curvature operator must be finite-dimensional, which generalizes work of Tam.

1. Introduction

Let (M^m, g) be an m-dimensional manifold with complete Riemannian metric g, where $m \geq 3$. We assume that the curvature operator of M is asymptotically nonnegative and we focus on the space of polynomial-growth harmonic p-forms of degree at most d on the manifold. Classical de Rham–Hodge theory implies that in the compact case the dimension of the space of harmonic forms is a topological invariant of the manifold, hence independent of the choice of the Riemannian metric. For complete noncompact manifolds, this topological invariance is no longer true. Nonetheless, it is an important question to study the space of harmonic forms and to seek topological and geometrical links. Yau [1975] proved that any positive harmonic function on a manifold with nonnegative Ricci curvature must be constant; hence the strong Liouville property holds. Saloff and Coste [1992] extended the result to the case where any Riemannian metric g' is uniformly equivalent to g. Thus, the space of positive harmonic functions is stable under a quasi-isometry for (M, g).

A complete manifold M is said to satisfy a Sobolev inequality $S(A, \nu)$ if there exist a point $q \in M$ and constants $A > 0$, $\nu > 2$, such that for all $r > 0$ and all $f \in C_0^\infty(B_q(r))$, we have

$$\int_{B_q(r)} |f|^{2\nu/(\nu-2)} \leq Ar^2 V(q, r)^{-2/\nu} \int_{B_q(r)} (|\nabla f|^2 + r^{-2} f^2),$$

MSC2000: primary 58A05; secondary 58A10.
Keywords: harmonic forms, curvature operator.
Sung was partially supported by the NSC..

where $V(q,r)$ is the volume of the geodesic ball $B_q(r)$. Examples include minimal submanifolds with Euclidean volume growth in $\mathbb{R}^m$ and manifolds with a nonnegative Ricci curvature. If a manifold satisfies a Sobolev inequality when endowed with a certain complete Riemannian metric, it obviously satisfies such an inequality (possibly with a different A) for any uniformly equivalent metric.

Li and Wang [1999], extending earlier work of Li [1997], proved that the dimension of the space $H_d^0(M)$ of polynomial-growth harmonic functions of growth order at most d has an estimate

$$\dim H_d^0(M) \le C(A,\nu)d^\nu$$

provided that the underlying manifold satisfies the Sobolev inequality $S(A,\nu)$. So the finite dimensionality of the space $H_d^0(M)$ is valid on such a manifold with respect to any uniformly equivalent metric.

Concerning general harmonic p-forms, Li [1997] established a dimension estimate of the space of polynomial-growth harmonic forms. Assuming that K_p, defined as the curvature operator on M if $p \ge 1$, is nonnegative, Li proved that

$$\dim H_d^p(M) \le Cd^{m-1},$$

where $H_d^p(M)$ denotes the space of polynomial-growth harmonic p-forms on M of growth order at most d. Recently, Chen and Sung [2006] showed that the stability of finite dimensionality of the space of $H_d^p(M)$ holds true under any uniformly equivalent metric on such manifold M.

Interestingly, Tam [1998] proved that if M is a complete manifold with nonnegative Ricci curvature outside a compact set, and if each unbounded component of $M\backslash D$, where D is a compact smooth domain in M, satisfies a certain kind of volume comparison property, then the space of polynomial-growth harmonic functions of degree at most d is finite dimensional. Furthermore, he proves the finite dimensionality of the space of polynomial-growth harmonic forms with a fixed growth rate on manifolds with asymptotically nonnegative curvature operator and the volume comparison property. The curvature operator K_p of M is asymptotically nonnegative if $K_p \ge -K(r)$, where $K(r) : [0,\infty) \to [0,\infty)$ is a nonnegative nonincreasing continuous function of distance r to a fixed point $q \in M$ which satisfies the integrability condition

$$\int_0^\infty rK(r) < \infty.$$

In view of the preceding results on the space $H_d^p(M)$, one would naturally to ask if the dimension of the space $H_d^p(M)$ is stable under a uniformly equivalent metric on M with asymptotically nonnegative curvature operator. An objective of this paper is to establish that the dimension of the space $H_d^p(M)$ remains finite under a uniformly equivalent metric on M.

It is proved in [Li and Tam 1992] that if the Ricci curvature of M is asymptotically nonnegative then M has finitely many ends. Also, by [Li and Tam 1995, Proposition 3.8], the volume comparison condition holds on M if M is a complete noncompact manifold with asymptotically nonnegative sectional curvature. However, it remains an open question whether an end of M will satisfy the volume comparison condition, if we only assume that M has nonnegative Ricci curvature outside a compact set.

Main Theorem 1.1. *Let (M^m, g) be a complete Riemannian manifold. Suppose that the curvature operator K_p is asymptotically nonnegative on (M, g) and the metric g' is uniformly equivalent to g on M. Then there exist constants $C > 0$ and $\nu > 2$ such that the dimension of the space $H_d^p(M, g')$ is finite and satisfies the inequality*

$$\dim H_d^p(M, g') \leq C\, d^\nu$$

for all $d \geq 1$, $p > 1$.

Remark 1.2. For $p = 1$, the curvature operator becomes Ricci curvature on M, we must assume that the first Betti number of M is finite so that the volume comparison condition holds true; compare [Li and Tam 1995]. Under this assumption, the theorem is valid.

An immediate consequence is that the space of polynomial-growth harmonic forms on the connected sum manifolds with nonnegative curvature operator must be finite-dimensional under quasi-isometry. Moreover,

Corollary 1.3. *Let (M, g) be a complete Riemannian manifold has nonnegative curvature operator outside a compact set, with finite first Betti number. If the metric g' is uniformly equivalent to g on M. Then there exist constants $C > 0$ and $\nu > 2$ such that the dimension of the space $H_d^p(M, g')$ is finite and satisfies the inequality*

$$\dim H_d^p(M, g') \leq C\, d^\nu$$

for all $d \geq 1$, $p \geq 1$.

We say g' is uniformly equivalent to g if there is some positive constant c such that $c^{-1}g' \leq g \leq cg'$ in the sense of bilinear forms. In other words, the Riemannian manifolds (M, g) and (M, g') are then called quasi-isometric. Clearly, quantities such as the distance and volume are uniformly equivalent under quasi-isometry; in particular, a Riemannian manifold quasi-isometric to a complete manifold is also complete. However, in general, any quantity involving derivatives of the Riemannian metric will not be comparable under a quasi-isometry; in particular, the same curvature condition is not expected to hold under a uniformly equivalent change of the metric. We overcome the difficulty by relating the space of harmonic forms to the eigenvalues of the Hodge Laplacian on the Busemann balls of the manifold with respect to the absolute boundary conditions. This idea was first introduced and successfully pursued in [Li and Wang 1999] for the harmonic functions. Here, our additional steps are to give the lower bound estimate of the eigenvalue of p form on Busemann balls of M, also show that such eigenvalues on each end of manifold are comparable under a uniformly equivalent change of the metric. We obtain a lower bound estimate of eigenvalue by modifying an argument from [Li 1980] in Section 2. Our main result is then proved in Section 3.

Throughout the paper, we assume that the first Betti number of M is finite for the case of $p = 1$.

2. Eigenvalue estimates

Let (M^m, g) be a complete, oriented Riemannian manifold with dimension m. The Hodge–Laplace–Beltrami operator Δ acting on the space of smooth p-forms $\Lambda^p(M)$ is defined as

$$\Delta = d\delta + \delta d,$$

here d denotes the exterior differential operator and $\delta = * d *$, where the linear operator $*$ is defined point-wise by

$$*(w_1 \wedge \cdots \wedge w_p) = w_{p+1} \wedge \cdots \wedge w_m$$

for a positively oriented orthonormal coframe $\{w_1,\ w_2, \ldots,\ w_m\}$ at the point. A p-form $w \in \Lambda^p(M)$ is called a harmonic p-form on (M, g) if

$$\Delta_g w = 0.$$

Let q denote a point on (M, g) and let $r_q(x)$ represent the geodesic distance function from $x \in M$ to the point q. For each $d \geq 0$, we denote the space of polynomial-growth harmonic p-forms of degree at most d by

$$H_d^p(M, g) \equiv \{w \in \Lambda^p(M) \mid \Delta_g w = 0,\ \text{and}\ |w| = O(r_q^d)\}.$$

For a bounded smooth domain $B \subset M$, a p-form w is said to satisfy the absolute condition on B if the tangential component of both w and δw on the boundary ∂B are zeros. On the boundary ∂B, let $N_{\partial B}$ (respectively $N^*_{\partial B_q}$) represent the inward unit normal vector (respectively covector) field. Now, denote exterior multiplication by $ext(\cdot)$ and dual exterior multiplication by $int(\cdot)$. It is not difficult to verify that Δ is a self-adjoint nonnegative operator on the space $\Lambda^p(B)$ of smooth p-forms on B satisfying the absolute boundary condition. By the standard elliptic theory, we see that Δ has a countable set of eigenvalues and the multiplicity of each eigenvalue is finite. If we list all the eigenvalues with multiplicity in nondecreasing order by

$\{\lambda_k,\ k = 1, 2, 3, \dots\}$, then $\lambda_k \to \infty$ as $k \to \infty$. Moreover, the i-th eigenvalue can be characterized as

$$\lambda_i = \inf_{\dim V = i} \sup_{w \in V \setminus \{0\}} R(w),$$

where V is a subspace of $\Lambda^p(B)$ and the Rayleigh–Ritz quotient $R(w)$ is defined by

$$R(w) = \frac{(dw, dw) + (\delta w, \delta w)}{(w, w)}$$

for $w \in \Lambda^p(B)$ and the L^2 inner product for two forms v and w in $\Lambda^p(B)$ is defined by

$$(v, w) = \int_B \langle v, w \rangle \, dx$$

with $\langle v, w \rangle$ being the point-wise inner product between v and w.

On the other hand, the Hodge-de Rham theorem provides an orthogonal decomposition of the space $\Lambda^p(B)$ of differential forms of degree p on B. For any $w \in \Lambda^p(B)$, w can be uniquely written as

$$w = h + dv + \delta u,$$

where $h \in H_p(B)$, the space of harmonic p-forms satisfying the absolute boundary condition and $v \in \Lambda^{p-1}(B)$, $u \in \Lambda^{p+1}(B)$. Clearly, the operator Δ leaves this decomposition invariant, and the eigenvalues of Δ on the subspace $H_p(B)$ are zeros.

Denote by $\{\mu_j^e(g) \mid j \geq 1\}$ the eigenvalues of Δ acting on the subspace $d\Lambda^{p-1}(B)$ of exact p-forms, and by $\{\mu_l^{co}(g) \mid l \geq 1\}$ those corresponding to the subspace $\delta\Lambda^{p+1}(B)$ of coexact p-forms. Then the eigenvalues $\{\lambda_i(g) \mid i > \dim H_p(B)\}$ is equal to the reordered union of $\{\mu_j^e(g) \mid j \geq 1\}$ and $\{\mu_l^{co}(g) \mid l \geq 1\}$. We have,

$$\{\lambda_i(g) \mid i > \dim H_p(B)\} = \{\mu_j^e(g) \mid j \geq 1\} \cup \{\mu_l^{co}(g) \mid l \geq 1\}.$$

The next lemma is essentially due to [Dodziuk 1982]; a proof is given in [Chen and Sung 2006].

Lemma 2.1. *Let (M, g) be a complete manifold with Riemannian metric g. Let g' denote another Riemannian metric on M which is uniformly equivalent to g. Then*

$$\dim H_p(B, g) = \dim H_p(B, g'),$$

and there exists a positive constant C such that

$$C^{-1}\lambda_i(g) \leq \lambda_i(g') \leq C\lambda_i(g)$$

for all $i \geq 1$.

We will obtain a lower bound estimate of the eigenvalues of p-forms satisfying the absolute boundary condition on Busemann balls of a manifold with asymptotically nonnegative curvature operator. The argument closely follows those in [Li 1980] and [Chen and Sung 2006].

An end E of a manifold M is an unbounded component of the complement of some compact smooth subset D of M. In this case, E is called an end corresponding to D. We say that M has finitely many ends if there exists $b < \infty$ such that the number of ends corresponding to D is less than or equal to b for any compact subset $D \subset M$. Let $E_1, E_2, \dots, E_L$ be the ends of M with respect to D. We say that E satisfies volume comparison property if there exists a constant $\zeta > 0$ such that

$$V_E(r) \le \zeta V_x\left(\frac{r}{2}\right), \tag{1}$$

for all $x \in \partial B_E(r)$ and for r large enough. Here we use $B_E(r)$ to denote $B_q(r) \cap E$, $\partial B_E(r) = \partial B_q(r) \cap E$, and $V_E(r)$ is the volume of $B_E(r)$. Also, denote

$$B_E(r_1, r) = B_E(r) \backslash B_E(r_1),$$

and

$$\partial B_E(r_1, r) = \partial B_E(r) \cup \partial B_E(r_1),$$

where $r_1 \le r$. By [Tam 1998, Lemma 1.1], the volume doubling property holds on ends $\{E_l\}_{l=1}^L$ of M, that is for $r > 2r_0$,

$$V_{E_l}((1+\varepsilon)r) \le (1+\varepsilon)^\mu \; V_{E_l}(r) \tag{2}$$

where $\mu > 0$ is a constant depending only on m. Moreover, given any $\eta > 0$, there is $r_1 > 2r_0$ such that for all $x \in \partial B_E(R)$, with $R > r_1$, and for all $\frac{3}{4}R > r' > r > 0$, we have

$$V_x(r') \le \left(\frac{r'}{r}\right)^{n+\eta} V_x(r). \tag{3}$$

Let $\gamma : [0, \infty) \to M$ be a ray with $\gamma(0) = q$, a fixed point in M; namely, γ is a geodesic of (M, g) and the geodesic distance $r(\gamma(t), \gamma(s))$ between $\gamma(t)$ and $\gamma(s)$ is equal to $|t - s|$ for all t and s in $[0, \infty)$. We define $b_t(x) = t - r(x, \gamma(t))$ for $t \ge 0$. For any fixed x, $b_t(x)$ is a nondecreasing function of t and $b_t(x) = r(q, \gamma(t)) - r(x, \gamma(t)) \le r(q, x)$. Therefore $b_\gamma(x) = \lim_{t\to\infty} b_t(x)$ exists for all $x \in M$. In fact, $b_t(x)$ converges uniformly on compact sets to $b_\gamma(x)$. Set

$$\beta(x) = \sup\{b_\gamma(x) \mid \gamma \text{ is a ray from } p\}.$$

Since for each x and for any ray γ, $b_\gamma(x) \le r(x, p)$, $\beta(x)$ is well defined and finite. We call $\beta(x)$ the Busemann function of (M, g) (based at point p).

We define $B(a) = \{x \mid \beta(x) \leq a\}$ as a Busemann ball on M, for some positive constant a. It is well known that $\beta(x)$ is proper and convex if the curvature operator of M is nonnegative. Therefore, the boundary of the Busemann ball is convex. Denote a Busemann ball on end E as $B_E(a) = B(a) \cap E$. The boundary of the Busemann ball $B_E(a)$ is convex if the curvature operator is nonnegative on each end. We will obtain a lower bound estimate of the eigenvalues of p-forms satisfying the absolute boundary condition on Busemann balls of a manifold with asymptotically nonnegative curvature operator. The argument closely follows those in [Li 1980] and [Chen and Sung 2006].

Lemma 2.2. *Let M^m be a complete manifold with asymptotically nonnegative curvature operator. If E is an end of M with respect to a compact subset $B_q(r_0) \subset M$, $r_0 > 0$, for a large enough $r > 2r_0$, such that $B_q(r_0)$ is contained in Busemann ball $B(r)$, then there exist constants $C > 0$ and $\nu > 2$ such that*

$$\dim H_p(B(r)) \leq C$$

and, for each $k > \dim H_p(B(r))$, there exists a constant C depending only on ν, m, p and η such that

$$\lambda_k(B(r)) \geq C k^{2/\nu} r^{-2}.$$

Proof. Let $\mathcal{V}$ be the k-dimensional space spanned by the eigen p-forms corresponding to the first k eigenvalues $\{\lambda_1, \ldots, \lambda_k\}$ on $B(r)$. Then there exists $w \in \mathcal{V}$, $w \neq 0$, such that

$$\frac{k}{V} \, \|w\|_2^2 \leq \|w\|_\infty^2 \cdot \min\{\tbinom{m}{p},\ k\}, \tag{4}$$

where $V = V\,(B(r))$ denotes the volume of $B(r)$. This is a result in [Li 1980].

On the other hand, we claim that there exist constants $C > 0$, $k_0 > 0$ and $\nu > 2$ such that for $w \in \mathcal{V}$,

$$\|w\|_\infty^2 \leq C V^{-1} r^{-\nu} \lambda_k^{\nu/2} \, \|w\|_2^2 \tag{5}$$

for all $k \geq k_0$. It is easy to see that the Lemma follows by combining inequalities (4) and (5). To prove (5), by the convexity of Busemann function and [Donnelly and Li 1982, Lemma 6.2], we first observe that for $w \in \Lambda^p(B(r))$,

$$\frac{\partial |w|^2}{\partial n} \leq 0$$

on $\partial B(r)$, where $\partial/\partial n$ is the outward unit normal of $\partial B(r)$. Since curvature operator K_p is asymptotically nonnegative, it means $K_p \geq -K(r)$, where $K(r)$: $[0, \infty) \to [0, \infty)$ is a nonnegative nonincreasing continuous function of distance

r to a fixed point $q \in M$ which satisfies the integrability condition

$$\int_0^\infty rK(r) < \infty.$$

By the argument in [Saloff-Coste 1992], we obtain a local weak Poincaré inequality on the Busemann ball $B(r)$. Also, using doubling volume condition and the local weak Poincaré inequality for the Busemann ball $B(r)$, we have the Sobolev–Poincaré inequality on $B(r)$ (see [Hajłasz and Koskela 1995])

$$\left(\int_{B(r)} |f - f_B|^{2\nu/(\nu-2)}\right)^{(\nu-2)/\nu} \le AV^{-2/\nu} r^2 \int_{B(r)} |\nabla f|^2,$$

where $\nu > 2$, $f_B = V^{-1}(B(r)) \int_{B(r)} f$, $A > 1$ is a constant depending only on m and η, and $V = V(B(r))$ is the volume of $B(r)$. Moreover, by [Li 1980], we observe that the Neumann Sobolev-type inequality

$$\left(\int_{B(r)} |f|^{2\nu/(\nu-2)}\right)^{(\nu-2)/\nu} \le AV^{-2/\nu} r^2 \left(\int_{B(r)} |\nabla f|^2 + r^{-2} \int_{B(r)} |f|^2\right) \tag{6}$$

holds on $B(r)$.

Let $\{w_i\}_{i=1}^k$ be the p-eigenforms satisfying the absolute boundary condition with the corresponding nonzero eigenvalues $\{\lambda_i\}_{i=1}^k$ and we also assume $\{w_i\}_{i=1}^k$ are orthonormal and span $\mathcal{V}$. If $w \in \mathcal{V}$, then there exist $\{a_i\}_{i=1}^k$ such that $w = \sum_{i=1}^k a_i w_i$, that is, $\Delta w = \sum_{i=1}^k \lambda_i a_i w_i$. By the Bochner formula,

$$\tfrac{1}{2}\Delta |w|^2 \le \langle \Delta w, w\rangle - |\nabla w|^2 + K|w|^2,$$

where K is the lower bound of curvature for all x in $B(r)$. Using the fact in [Li 1980, Lemma 8],

$$|\nabla |w||^2 \le |\nabla w|^2.$$

Thus,

$$\tfrac{1}{2}\Delta |w|^2 \le \langle \Delta w, w\rangle - |\nabla |w||^2 + K|w|^2. \tag{7}$$

Let $\alpha \ge 1$, and we multiply both sides of this inequality ((7)) by $|w|^{2\alpha-2}$ and integrate over $B(r)$,

$$\begin{aligned} \frac{1}{2}\int_{B(r)} |w|^{2\alpha-2} \Delta |w|^2 \\ \le \int_{B(r)} |w|^{2\alpha-2} \langle \Delta w, w\rangle + \int_{B(r)} K|w|^{2\alpha} - \int_{B(r)} |w|^{2\alpha-2} |\nabla |w||^2. \end{aligned} \tag{8}$$

Using the absolute boundary condition, the left-hand side of this inequality becomes

$$\frac{1}{2}\int_{B(r)} |w|^{2\alpha-2}\,\Delta|w|^2 \geq (\alpha-1)\int_{B(r)} |w|^{2\alpha-3}\left\langle \nabla|w|, \nabla|w|^2\right\rangle$$
$$= \frac{2(\alpha-1)}{\alpha^2}\int_{B(r)} \left\langle \nabla|w|^{\alpha}, \nabla|w|^{\alpha}\right\rangle,$$

and the third term in the right-hand side of (8) can be rewritten as

$$\int_{B(r)} |w|^{2\alpha-2}\,|\nabla|w||^2 = \frac{1}{\alpha^2}\int_{B(r)} \left|\nabla|w|^{\alpha}\right|^2.$$

Hence, (8) becomes

$$\frac{2(\alpha-1)}{\alpha^2}\int_{B(r)} \left|\nabla|w|^{\alpha}\right|^2 \leq \int_{B(r)} |w|^{2\alpha-2}\langle\Delta w, w\rangle + \int_{B(r)} K|w|^{2\alpha} - \frac{1}{\alpha^2}\int_{B(r)}\left|\nabla|w|^{\alpha}\right|^2. \tag{9}$$

With $f = |w|$, this can be rewritten as

$$\frac{2\alpha-1}{\alpha^2}\int_{B(r)} \left|\nabla f^{\alpha}\right|^2 \leq \int_{B(r)} f^{2\alpha-2}\langle\Delta w, w\rangle + \int_{B(r)} Kf^{2\alpha}. \tag{10}$$

Applying Neumann Sobolev-type inequality (6) to the function f^{α}, one has

$$\left(\int_{B(r)} |f|^{2\beta\alpha}\right)^{1/\beta} \leq AV^{-2/\nu}r^2\left(\int_{B(r)}\left|\nabla f^{\alpha}\right|^2 + r^{-2}\int_{B(r)} f^{2\alpha}\right), \tag{11}$$

where $\beta = \frac{\nu}{\nu-2}$. Thus (10) and (11) suggest

$$\|f\|_{2\alpha\beta}^{2\alpha} \leq \frac{\alpha^2}{2\alpha-1}AV^{-2/\nu}r^2\int_{B(r)} f^{2\alpha-2}\langle\Delta w, w\rangle + AV^{-2/\nu}r^2\left(\frac{\alpha^2}{2\alpha-1}\int_{B(r)} Kf^{2\alpha} + r^{-2}\|f\|_{2\alpha}^{2\alpha}\right). \tag{12}$$

By the Hölder inequality, we have

$$\int Kf^{2\alpha} \leq BV^{1/q}\left(\int \left(f^{2\alpha}\right)^{q/(q-1)}\right)^{(q-1)/q}$$
$$\leq BV^{1/q}\left(\int f^{2\alpha}\right)^{\frac{\beta(q-1)-q}{q(\beta-1)}}\left(\int f^{2\alpha\beta}\right)^{\frac{1}{q(\beta-1)}},$$

where $q > \dfrac{\beta}{\beta-1} = \dfrac{\nu}{2}$ and $B = \left(V^{-1}\int_{B(r)} |K|^q\right)^{1/q}$. However, applying the inequality

$$h^{\varepsilon} \le \delta_1^{(\varepsilon-1)/\varepsilon} h + \delta_1 \varepsilon^{1/(1-\varepsilon)}\left(\frac{1}{\varepsilon}-1\right)$$

by setting $\varepsilon = \dfrac{\beta\,(q-1)-q}{q\,(\beta-1)}$ and

$$h = \left(\frac{\alpha^2}{2\alpha-1} ABV^{(\nu-2q)/q\nu} r^2\right)^{\frac{q(\beta-1)}{\beta(q-1)-q}} \left(\int f^{2\alpha}\right)\left(\int f^{2\alpha\beta}\right)^{\frac{-1}{\beta}}.$$

Then

$$\begin{aligned}\frac{\alpha^2}{2\alpha-1} & ABV^{(\nu-2q)/q\nu} r^2 \left(\int f^{2\alpha}\right)^{\frac{\beta(q-1)-q}{q(\beta-1)}} \left(\int f^{2\alpha\beta}\right)^{\frac{q-\beta(q-1)}{q\beta(\beta-1)}} \\ &\le \delta_1^{(\varepsilon-1)/\varepsilon}\left(\frac{\alpha^2}{2\alpha-1} ABV^{(\nu-2q)/q\nu} r^2\right)^{\frac{2q}{2q-\nu}} \left(\int f^{2\alpha}\right)\left(\int f^{2\alpha\beta}\right)^{\frac{-1}{\beta}} \\ &\qquad + \delta_1 \varepsilon^{1/(1-\varepsilon)}\left(\frac{1}{\varepsilon}-1\right).\end{aligned}$$

If we select δ_1 small enough, and since $\dfrac{\alpha^2}{2\alpha-1} \ge 1$, Equation (12) can be rewritten as

$$\begin{aligned}\|f\|_{2\alpha\beta}^{2\alpha} &\le \frac{1}{1-\delta_1 c\,(\varepsilon)}\left(\frac{\alpha^2}{2\alpha-1}\right)^{\frac{2q}{2q-\nu}} AV^{-2/\nu} r^2 \\ &\times \left(\int_{B(r)} f^{2\alpha-2}\,\langle \Delta w, w\rangle + \left(\delta_1^{\nu/(\nu-2q)} B^{2q/(2q-\nu)} (Ar^2)^{\nu/(2q-\nu)} + \frac{1}{r^2}\right)\|f\|_{2\alpha}^{2\alpha}\right).\end{aligned}$$

Let $\alpha = \beta^i,\ i = 0, 1, 2, \ldots$. Then

$$\|f\|_{2\beta^{i+1}}^{2\beta^i} \le \tilde{C}\left(\int_{B(r)} f^{2\beta^i-2}\,\langle \Delta w, w\rangle + \left(\bar{K}+r^{-2}\right)\|f\|_{2\beta^i}^{2\beta^i}\right). \tag{13}$$

where

$$\tilde{C} = \frac{1}{1-\delta_1 c(\varepsilon)}\left(\frac{\alpha^2}{2\alpha-1}\right)^{\frac{2q}{2q-\nu}} AV^{-2/\nu} r^2 \quad \text{and} \quad \bar{K} = B^{2q/(2q-\nu)}\left(\delta_1^{-1} Ar^2\right)^{\nu/(2q-\nu)}.$$

When $i = 0$, (13) gives

$$\|f\|_{2\beta}^2 \le \frac{1}{1-\delta_1 c\,(\varepsilon)} AV^{-2/\nu} r^2 \left(\int_{B(r)} \langle \Delta w, w\rangle + \left(\bar{K}+r^{-2}\right)\|f\|_2^2\right).$$

Since

$$\int_{B(r)} \langle \Delta w, w\rangle = \int_{B(r)} \langle \lambda_i a_i w_i, a_j w_j\rangle = \lambda_i a_i^2 \le \lambda_k a_i^2 = \lambda_k \int_{B(r)} \langle w, w\rangle,$$

this implies

$$\|f\|_{2\beta}^2 \le \frac{1}{1-\delta_1 c(\varepsilon)} AV^{-2/\nu} r^2 (\lambda_k + \bar{K} + r^{-2}) \|f\|_2^2 .$$

By the Hölder inequality,

$$\|f\|_2^2 \le V^{(\beta-1)/\beta} \|f\|_{2\beta}^2 .$$

We conclude that

$$V^{-(\beta-1)/\beta} \|f\|_2^2 \le \frac{1}{1-\delta_1 c(\varepsilon)} AV^{-2/\nu} r^2 \left(\lambda_k + \bar{K} + r^{-2}\right) \|f\|_2^2 .$$

Now we claim that for $1 \le i < \infty$,

$$(14) \quad V^{-(\beta-1)/\alpha\beta} \|f\|_{2\alpha}^2 \le \prod_{j=0}^{i} \left(\frac{\beta^{2j}}{2\beta^j - 1}\right)^{\frac{2q}{2q-\nu}\beta^{-j}} \left(\frac{1}{1-\delta_1 c(\varepsilon)} AV^{-2/\nu} r^2 \lambda_k^*\right)^{\sum_{j=0}^{i} \beta^{-j}} \|f\|_2^2 ,$$

where $\lambda_k^* = \lambda_k + \bar{K} + r^{-2}$. Assuming this inequality (14) is true for $\alpha = \beta^j$, $j = 0, \ldots, i-1$, by induction, we need to show that (14) is still valid for $j = i$. Suppose $g = |\bar{w}|$, where $\bar{w} \in \mathcal{V}$ with the property that

$$(15) \qquad \frac{\|g\|_{2\alpha}}{\|g\|_2} \ge \frac{\|w\|_{2\alpha}}{\|w\|_2} \quad \text{for all } w \in \mathcal{V}.$$

Without loss of the generality, we may use the scaling and assume $\|g\|_2 = 1$. By the Hölder inequality, Equation (13) implies

$$(16) \qquad \|g\|_{2\alpha\beta}^{2\alpha} \le \tilde{C} \left(\int_{B(r)} g^{2\alpha-2} \langle \Delta\bar{w}, \bar{w}\rangle + \left(\bar{K} + r^{-2}\right) \|g\|_{2\alpha}^{2\alpha} \right)$$
$$\le \tilde{C}\left(\|g\|_{2\alpha}^{2\alpha-1} \|\Delta\bar{w}\|_{2\alpha} + \left(\bar{K} + r^{-2}\right) \|g\|_{2\alpha}^{2\alpha}\right).$$

We also note that if $s \ge 2$, then there exists a subset $\{\sigma\} \subset \{1, 2, \ldots, k\}$ such that

$$\left\| \sum_{i=1}^{k} \lambda_i w_i \right\|_s \le \left\| \sum_{\sigma} \lambda_k w_\sigma \right\|_s .$$

This is proved in [Li 1980, Lemma 17]. Hence, let $\bar{w} = \sum b_i w_i$, then $\Delta\bar{w} = \sum \lambda_i b_i w_i$ and we have

$$\|\Delta\bar{w}\|_{2\alpha} = \left\| \sum_i \lambda_i b_i w_i \right\|_{2\alpha} \le \left\| \sum_\sigma \lambda_k b_\sigma w_\sigma \right\|_{2\alpha} = \lambda_k \left\| \sum_\sigma b_\sigma w_\sigma \right\|_{2\alpha}$$
$$\le \lambda_k \|g\|_{2\alpha} \left\| \sum_\sigma b_\sigma w_\sigma \right\|_2 \quad \text{by (15)}$$
$$\le \lambda_k \|g\|_{2\alpha} .$$

It is obvious that (16) gives

$$\|g\|_{2\alpha\beta}^{2\alpha} \le \tilde{C}\lambda_k^* \|g\|_{2\alpha}^{2\alpha},$$

where $\lambda_k^* = \lambda_k + \bar{K} + r^{-2}$. By the method of iteration, we obtain

$$\|g\|_{2\alpha\beta}^2 \le \tilde{C}_i \, (\lambda_k^*)^{\sum_{j=0}^i \beta^{-j}} \|g\|_2^2,$$

where $\alpha = \beta^i$ and

$$\tilde{C}_i = \prod_{j=0}^i \left(\frac{\beta^{2j}}{2\beta^j - 1}\right)^{2q/(2q-\nu)\beta^{-j}} \left(\frac{1}{1-\delta_1 c\,(\varepsilon)} A V^{-2/\nu} r^2\right)^{\sum_{j=0}^i \beta^{-j}}.$$

On the other hand, by Hölder inequality, we have

$$\|g\|_{2\alpha}^2 \le V^{(\beta-1)/\alpha\beta} \|g\|_{2\alpha\beta}^2.$$

Therefore,

$$V^{-(\beta-1)/\alpha\beta} \|g\|_{2\alpha}^2 \le \tilde{C}_i \cdot \left(\lambda_k^*\right)^{\sum_{j=0}^i \beta^{-j}} \|g\|_2^2. \tag{17}$$

Applying (15) to (17), it is easy to check that

$$V^{-(\beta-1)/\alpha\beta} \|f\|_{2\beta^i}^2 \le \tilde{C}_i \cdot (\lambda_k^*)^{\sum_{j=0}^i \beta^{-j}} \|f\|_2^2.$$

Letting $i \to \infty$, due to $\sum_{j=0}^\infty \beta^{-j} = \nu/2$ and

$$\prod_{j=0}^\infty \left(\frac{\beta^{2j}}{2\beta^j - 1}\right)^{\beta^{-j}} \le \exp \frac{1}{\beta^{1/2} - 1} = c_1(\nu),$$

we conclude that

$$\|f\|_\infty^2 \le C(\nu, q) \left(A V^{-2/\nu} r^2 \lambda_k^*\right)^{\nu/2} \|f\|_2^2,$$

where A is a positive constant depending only on m and η. This means, for all $w \in \mathcal{V}$, w satisfies

$$\|w\|_\infty^2 \le C(\nu, \eta, m) V^{-1} r^\nu \left(\lambda_k^*\right)^{\nu/2} \|w\|_2^2,$$

for all $q > \nu/2$, where

$$\lambda_k^* = \lambda_k + \bar{K} + r^{-2} \quad \text{and} \quad \bar{K} = \delta_1^{\nu/(\nu-2q)} B^{2q/(2q-\nu)} \left(Ar^2\right)^{\nu/(2q-\nu)}.$$

In fact, by the assumption of curvature, we have $B \le 1$; hence

$$\|w\|_\infty^2 \le C(\nu, \eta, m) V^{-1} r^\nu \left(\lambda_k + c(\delta_1, \nu) r^{2\nu/(2q-\nu)} + r^{-2}\right)^{\nu/2} \|w\|_2^2. \tag{18}$$

We note that the Hodge Laplace Beltrami operator $\Delta = d\delta + \delta d$ is nonnegative and self-adjoint on $B(r)$ under the absolute boundary condition. Hence, using the standard elliptic theory, if we assume λ_1 be the first nonzero eigenvalue, then we have

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \to \infty.$$

This means, there exists k_0 large enough such that the k-th nonzero eigenvalue

$$\lambda_k \geq \max\{c(\delta_1, \nu) r^{2\nu/(2q-\nu)}, r^{-2}\}. \tag{19}$$

Thus, by (18) and (19), we have

$$\|w\|_\infty^2 \leq C(\nu, \eta, m) V^{-1} r^\nu \lambda_k^{\nu/2} \|w\|_2^2,$$

for all w belong to the space $\mathcal{V}$ spanned by the p-eigenforms corresponding to the first k nonzero eigenvalues. Therefore, the dimension estimate (4) gives

$$\frac{k}{V} \|w\|_2^2 \leq \binom{m}{p} \|w\|_\infty^2 \leq C(\nu, \eta, m, p) V^{-1} r^\nu \lambda_k^{\nu/2} \|w\|_2^2,$$

and we conclude that

$$\lambda_k(B(r)) \geq C k^{2/\nu} r^{-2}$$

for all $k \geq k_0$, where $C = C\,(\nu, \eta, m, p)$ is a positive constant. □

3. Main Result

Let (M^m, g), $m \geq 3$, be a complete noncompact manifold with Riemannian metric g. We consider the manifold M with its curvature operator $K_p(x)$ is asymptotically nonnegative. By [Li and Tam 1992], we know that M has finitely many ends if the curvature operator $K_p(x)$ of M is asymptotically nonnegative. Assume $E_1, \ldots, E_L$ be the ends of M with respect to a compact smooth domain $B_q(r_0)$ in M. Let $\mathfrak{B}$ be a n-dimensional vector bundle over M with a metric. For $r > 4r_0 > 0$, with $B_q\,(r_0) \subset B(r)$, where $B(r)$ is a Busemann ball in M . We define a positive semidefinite symmetric bilinear form S_r on the space of section $\Gamma\,(\mathfrak{B})$ of $\mathfrak{B}$ by

$$S_r\,(u, v) = V^{-1}(r) \int_{B(r)} \langle u, v\rangle \tag{20}$$

for u, $v \in \Gamma\,(\mathfrak{B})$. In particular, S_r is always positive definite, and $(\mathfrak{B}, S_r)$ is an inner product space in $B(r)$.

Suppose each end satisfies volume comparison property, the volume doubling property holds on ends $\{E_l\}_{l=1}^L$ of M, that is for $r > 2r_0$ and $\varepsilon > 0$,

$$V_{E_l}\,((1+\varepsilon)\,r) \leq (1+\varepsilon)^\mu\, V_{E_l}(r),$$

where $\mu > 0$ is a constant depending only on m. Moreover, given any $\eta > 0$, there is $r_1 > 2r_0$ such that for all $x \in \partial B_E\,(R)$, with $R > r_1$, and for all $\frac{3}{4}R > r' > r > 0$, we have

$$V_x\left(r'\right) \le \left(\frac{r'}{r}\right)^{n+\eta} V_x(r).$$

The curvature operator K_p on M is defined by

$$K_p \ge -K = \begin{cases} \text{lower bound of curvature operator} & \text{if } p > 1, \\ \text{lower bound of Ricci curvature} & \text{if } p = 1. \end{cases}$$

Concerning the asymptotically nonnegative curvature operator, the volume comparison property holds on each end for $p > 1$ [Li and Tam 1995, Proposition 3.8]. For the case of $p = 1$, the curvature operator is the Ricci curvature, the volume comparison property holds on ends of M if we assume that the first Betti number of M is finite.

Lemma 3.1. *Let $\mathcal{V}$ be a k-dimensional subspace of a vector space W. Assume that W is endowed with an inner production L and a bilinear form Φ. Then for any given linearly independent set of vectors $\{w_1, \dots, w_{k-1}\} \subset W$, there exists an orthonormal basis $\{v_1, \dots, v_k\}$ of $\mathcal{V}$ with respect to L such that $\Phi(v_i, w_j) = 0$ for all $1 \le j < i \le k$.*

Lemma 3.2. *Let M be a complete Riemannian manifold with asymptotically nonnegative curvature operator. Let $\mathcal{V}$ be a k-dimensional subspace of $H_d^p(M, g)$, and let $E_1, E_2, \ldots, E_L$ be the ends of M with respect to $B_q(r_0)$, $r_0 > 0$. For any fixed $0 < \varepsilon < \frac{1}{4}$, $r > 4r_0$ and any subspace Y of $\mathcal{V}$, if $\{v_{s+1}, \dots, v_k\}$ is an orthonormal basis of inner production $S_{(1+\varepsilon)r}$ on Y. Then*

$$\sum_{i=s+1}^{k} S_r(v_i, v_i) \le \frac{8(1+\varepsilon)^{\mu}}{\varepsilon^2 r^2} \sum_{i=s+1}^{k} \lambda_i^{-1}(B((1+\varepsilon)r)),$$

where $\mu > 0$ is a constant depending only on m.

Proof. Let $\lambda_i\,(B((1+\varepsilon)r))$ denote the i-th nonzero eigenvalue of p-forms on Busemann ball $B((1+\varepsilon)r)$ satisfying the absolute boundary condition on $\partial B\,((1+\varepsilon)\,r)$. Let ϕ be a nonnegative function defined on $B((1+\varepsilon)r)$ satisfying these conditions:

$$\begin{aligned} \phi &= 1 \quad \text{on } B(r), \\ 0 \le \phi &\le 1 \quad \text{on } B((1+\varepsilon)r), \\ \phi &= 0 \quad \text{on } \partial B((1+\varepsilon)r), \end{aligned}$$

and

$$|\nabla\phi| \le \frac{2}{\varepsilon r}.$$

Observing that by the property of unique continuation, $\mathcal{V}$ is a k-dimensional subspace because

$$\mathcal{V} \subset L^2\left(B((1+\varepsilon)r), \phi dv\right) \cap L^2\left(B(r), dv\right).$$

Applying Lemma 3.1 with $\{w_1, \ldots, w_k\}$ as the eigen p-forms of Busemann ball $B((1+\varepsilon)r)$ corresponding to the nonzero eigenvalues

$$\left\{\lambda_1\left(B((1+\varepsilon)r)\right),\ \ldots,\ \lambda_k\left(B((1+\varepsilon)r)\right)\right\},$$

we get an orthonormal basis $\{v_1, \ldots, v_k\}$ of $\mathcal{V}$ with respect to the inner product $S_{(1+\varepsilon)r}$ satisfying

$$S_{(1+\varepsilon)r}\left(v_i, v_j\right) = V^{-1}\left((1+\varepsilon)\, r\right) \int_{B((1+\varepsilon)r)} \left\langle v_i, v_j\right\rangle.$$

Hence

$$\Phi(v_i, w_j) = \int_{B((1+\varepsilon)r)} \left\langle v_i, w_j\right\rangle \phi\, dv = 0$$

for $1 \le j < i \le k$. Thus, for any $1 \le i \le k$, let $|v_i|^2 = \langle v_i, v_i\rangle$, $\|v_i\|^2 = (v_i, v_i) = \int \langle v_i, v_i\rangle$ and $\mathrm{sgn} = (-1)^{m(p+1)+1}$. We have

$$\lambda_i\left(B((1+\varepsilon)r)\right) \int_{B((1+\varepsilon)r)} |\phi v_i|^2\, dv \le \|d(\phi v_i)\|^2 + \|\delta(\phi v_i)\|^2. \tag{21}$$

The right-hand side of this inequality can be rewritten as

$$\begin{aligned}(d(\phi v_i), d(\phi v_i)) &+ (\delta(\phi v_i), \delta(\phi v_i)) \\ &= (d\phi \wedge v_i + \phi d v_i, d\phi \wedge v_i + \phi d v_i) \\ &\qquad + \left(\phi\delta v_i + \mathrm{sgn}\,(d\phi \wedge * v_i)\,, \phi\delta v_i + \mathrm{sgn} * (d\phi \wedge * v_i)\right) \\ &= \|d\phi \wedge v_i\|^2 + 2(\phi d v_i, d\phi \wedge v_i) + \|\phi d v_i\|^2 + \|\phi\delta v_i\|^2 \\ &\qquad + 2\left(\phi\delta v_i, \mathrm{sgn} * (d\phi \wedge * v_i)\right) + \|d\phi \wedge * v_i\|^2.\end{aligned} \tag{22}$$

On the other hand,

$$\begin{aligned}0 &= \int_{B((1+\varepsilon)r)} \phi^2 \langle v_i, \Delta v_i\rangle\, dv = \left(\phi^2 v_i, \Delta v_i\right) = \left(\delta(\phi^2 v_i), \delta v_i\right) + \left(d(\phi^2 v_i), d v_i\right) \\ &= (\phi\delta v_i, \phi\delta v_i) + 2\left(\phi\delta v_i, \mathrm{sgn} * (d\phi \wedge * v_i)\right) + (\phi d v_i, \phi d v_i) + 2\left(\phi d v_i, d\phi \wedge v_i\right).\end{aligned}$$

Then (22) gives

$$\begin{aligned}\left(d(\phi v_i), d(\phi v_i)\right) + \left(\delta(\phi v_i), \delta(\phi v_i)\right) &= \|d\phi \wedge v_i\|^2 + \|d\phi \wedge * v_i\|^2 \\ &\le 2 \sup\nolimits_{B((1+\varepsilon)r)} |\nabla\phi|^2 \cdot \|v_i\|^2 \\ &\le \frac{8}{\varepsilon^2 r^2} V\left((1+\varepsilon)\, r\right),\end{aligned} \tag{23}$$

since v_i is orthonormal on $B((1+\varepsilon)r)$. Therefore, (21) and (23) imply

$$\begin{aligned}
\int_{B(r)} |v_i|^2\, dv &\le \int_{B((1+\varepsilon)r)} |\phi v_i|^2\, dv \\
&\le \lambda_i^{-1}(B((1+\varepsilon)r)) \left\{ \|d(\phi v_i)\|^2 + \|\delta(\phi v_i)\|^2 \right\} \\
&\le \frac{8}{\varepsilon^2 r^2 \lambda_i (B((1+\varepsilon)r))} V\left((1+\varepsilon)\,r\right).
\end{aligned} \tag{24}$$

Hence, if we let Y represent the space spanned by $\{v_{s+1}, \dots, v_k\}$, we get

$$\dim Y = k - s$$

and

$$\sum_{i=s+1}^{k} \int_{B(r)} |v_i|^2\, dv \le \sum_{i=s+1}^{k} \frac{8}{\varepsilon^2 r^2 \lambda_i (B((1+\varepsilon)r))} V\left((1+\varepsilon)\,r\right).$$

Therefore,

$$\sum_{i=s+1}^{k} V^{-1}(r) \int_{B(r)} |v_i|^2\, dv \le \sum_{i=s+1}^{k} \frac{8}{\varepsilon^2 r^2 \lambda_i (B((1+\varepsilon)r))} \frac{V\left((1+\varepsilon)\,r\right)}{V(r)}.$$

Moreover, volume doubling property holds on each end of M which implies

$$V\left((1+\varepsilon)\,r\right) \le (1+\varepsilon)^{\mu}\;\; V(r)$$

for $r > 2r_0$, where $\mu > 0$ is a constant depending only on m. We conclude

$$\sum_{i=s+1}^{k} S_r\,(v_i, v_i) \le \frac{8\,(1+\varepsilon)^{\mu}}{\varepsilon^2 r^2} \sum_{i=s+1}^{k} \lambda_i^{-1}(B((1+\varepsilon)r)). \qquad \square$$

Lemma 3.3. *Let $E_1, \dots, E_L$ be the ends of M with respect to $B\,(r_0)$, $r_0 > 0$, and let $\mathcal{V}$ be a k-dimensional vector space with polynomial growth of degree at most d. Then for all $0 < \varepsilon < \frac{1}{4}$ and $r_1 > 4r_0$, there is $r > r_1$ such that if $\{u_1, \dots, u_k\}$ is an orthonormal basis for $\mathcal{V}$ with respect to $S_{(1+\varepsilon)r}$, then*

$$\sum_{i=1}^{k} S_r\,(u_i, u_i) \ge k\,(1+\varepsilon)^{-(2d+1)}.$$

Proof. Denote the trace of S_r with respect to $S_{(1+\varepsilon)r}$ by $\mathrm{tr}_{(1+\varepsilon)r} S_r$. and the determinant of S_r with respect to $S_{(1+\varepsilon)r}$ by $\det_{(1+\varepsilon)r} S_r$. Suppose the lemma is false. Then, for some $0 < \varepsilon < \frac{1}{4}$ and $r_1 > 4r_0$ such that for all $r > r_1$,

$$\mathrm{tr}_{(1+\varepsilon)r} S_r = \mathrm{tr}_{(1+\varepsilon)r} S_r < k\,(1+\varepsilon)^{-(2d+1)}.$$

On the other hand, the arithmetic geometric mean asserts that

$$(\det_{(1+\varepsilon)r} S_r)^{1/k} \le k^{-1} \mathrm{tr}_{(1+\varepsilon)r} S_r.$$

This implies that

$$\det_{(1+\varepsilon)r} S_r \le (1+\varepsilon)^{-k(2d+1)}.$$

Setting $r = r_1 + 1$ and iterating this inequality j time, we obtain

$$\det_{(1+\varepsilon)^j r} S_r \le (1+\varepsilon)^{-jk(2d+1)}. \tag{25}$$

However, for a fixed S_r orthonormal basis $\{u_i\}_{i=1}^k$ of $\mathcal{V}$, and the polynomial growth assumption imply that there exists a constant $C > 0$, depending on $\mathcal{V}$, such that

$$S_{(1+\varepsilon)^j r}(u_i, u_i) = V^{-1}\left((1+\varepsilon)^j r\right) \int_{B\left((1+\varepsilon)^j r\right)} \langle u_i, u_i \rangle \le C\left((1+\varepsilon)^j r\right)^{2d}$$

for all $1 \le i \le k$. Hence

$$\det_r S_{(1+\varepsilon)^j r} \le k! C^k \left((1+\varepsilon)^j r\right)^{2dk}.$$

This contradicts (25) since $j \to \infty$. □

We are now ready to prove the Main Theorem, which we restate here.

Main Theorem. *Let (M^m, g), $m \ge 3$, be a complete noncompact manifold with metric g, and $q \in M$ be a fixed point. Suppose curvature operator K_p is asymptotically nonnegative. Let $E_1, E_2, \ldots, E_L$ be the ends of M with respect to $B_q(r_0)$, $r_0 > 0$. Then for any uniformly equivalent metric g' on M and for all $d \ge 1$, the space $H_d^p(M, g')$, is finite-dimensional and its dimension satisfies the inequality*

$$\dim H_d^p(M, g') \le C\, d^\nu$$

for some constants $\nu > 2$, $\mu > 0$, $\eta > 0$ and $C = C(m, p, \nu, \mu, \eta) > 0$.

Proof. For any k-dimensional subspace $\mathcal{V}$ of $H_d^p(M)$. By Lemma 3.3, if we set $\varepsilon = 1/5d$, there exists $r > 4r_0$ such that

$$\operatorname{tr}_{(1+\varepsilon)r} S_r \ge k\,(1+\varepsilon)^{-(2d+1)}.$$

Let λ_k be the k-th eigenvalue of the Hodge Laplacian acting on p-forms on Busemann ball $B((1+\varepsilon)\, r) \supset B(r_0)$ satisfying the absolute boundary condition on $\partial B((1+\varepsilon)r)$ under the metric g'. Then by Lemma 2.1, Lemma 2.2 and the assumption of K_p, we have

$$\lambda_k \ge C\, k^{2/\nu}\,(1+\varepsilon)^{-2}\, r^{-2},$$

for $k > s = \sum_{l=1}^{L} \dim H_p\,(B((1+\varepsilon)\, r))$, where C is a positive constant depending only on m, p, ν and η. Combining with Lemma 3.2, we find there exists a subspace Y in $\mathcal{V}$ with

$$\dim Y = \dim \mathcal{V} - \sum_{l=1}^{L} \dim H_p(B((1+\varepsilon)\, r)),$$

and a positive constant μ such that

$$\sum_{i=s+k_0+1}^{k} S_r(v_i, v_i) \le \frac{8(1+\varepsilon)^{\mu}}{\varepsilon^2 r^2} \sum_{i=s+1}^{k} \lambda_i^{-1}(B((1+\varepsilon)r)) \le \frac{8(1+\varepsilon)^{\mu+2}}{\varepsilon^2} k^{1-2/\nu},$$

where $\varepsilon = 1/5d$. Hence

$$\begin{aligned} k(1+\varepsilon)^{-(2d+1)} &\le \mathrm{tr}_{(1+\varepsilon)r} S_r(V) \le \mathrm{tr}_{(1+\varepsilon)r} S_r(Y) + \dim H_p(B((1+\varepsilon)r)) \\ &\le C\, d^2\, k^{1-2/\nu}. \end{aligned}$$

where we have used Lemmas 2.1 and 2.2. Therefore $k \le C\, d^{\nu}$. Since $\mathcal{V}$ is arbitrary, we conclude that

$$\dim H_d^p(M, g') \le C\, d^{\nu}$$

for all $d \ge 1$. □

Suppose M has nonnegative curvature operator outside a compact set, with finite first Betti number, In this case, even thought it is not exactly true that each end of M satisfies volume comparison property for $p = 1$, however, it is almost true so that by modifying some arguments of main theorem still holds for such a manifold; see [Li and Tam 1995, Corollary 6.2] in particular. Hence we have Corollary 1.3.

Acknowledgement

We thank Professor Jiaping Wang for his interest in this work.

References

[Chen and Sung 2006] J.-T. R. Chen and C.-J. A. Sung, “Dimension estimate of polynomial growth harmonic forms”, *J. Differential Geom.* **73**:1 (2006), 167–183. MR 2007b:53158 Zbl 05042956

[Dodziuk 1982] J. Dodziuk, “Eigenvalues of the Laplacian on forms”, *Proc. Amer. Math. Soc.* **85**:3 (1982), 437–443. MR 84k:58223 Zbl 0502.58038

[Donnelly and Li 1982] H. Donnelly and P. Li, “Lower bounds for the eigenvalues of Riemannian manifolds”, *Michigan Math. J.* **29**:2 (1982), 149–161. MR 83g:58069 Zbl 0488.58022

[Hajłasz and Koskela 1995] P. Hajłasz and P. Koskela, “Sobolev meets Poincaré”, *C. R. Acad. Sci. Paris Sér. I Math.* **320**:10 (1995), 1211–1215. MR 96f:46062 Zbl 0837.46024

[Li 1980] P. Li, “On the Sobolev constant and the p-spectrum of a compact Riemannian manifold”, *Ann. Sci. École Norm. Sup.* (4) **13**:4 (1980), 451–468. MR 82h:58054 Zbl 0466.53023

[Li 1997] P. Li, “Harmonic sections of polynomial growth”, *Math. Res. Lett.* **4**:1 (1997), 35–44. MR 98i:53054 Zbl 0880.53039

[Li and Tam 1992] P. Li and L.-F. Tam, “Harmonic functions and the structure of complete manifolds”, *J. Differential Geom.* **35**:2 (1992), 359–383. MR 93b:53033 Zbl 0768.53018

[Li and Tam 1995] P. Li and L.-F. Tam, “Green’s functions, harmonic functions, and volume comparison”, *J. Differential Geom.* **41**:2 (1995), 277–318. MR 96f:53054 Zbl 0827.53033

[Li and Wang 1999] P. Li and J. Wang, “Counting massive sets and dimensions of harmonic functions”, *J. Differential Geom.* **53**:2 (1999), 237–278. MR 2001k:53063 Zbl 1039.53036

[Saloff-Coste 1992] L. Saloff-Coste, "Uniformly elliptic operators on Riemannian manifolds", *J. Differential Geom.* **36**:2 (1992), 417–450. MR 93m:58122 Zbl 0735.58032

[Tam 1998] L.-F. Tam, "A note on harmonic forms on complete manifolds", *Proc. Amer. Math. Soc.* **126**:10 (1998), 3097–3108. MR 99a:53047 Zbl 0922.58017

[Yau 1975] S. T. Yau, "Harmonic functions on complete Riemannian manifolds", *Comm. Pure Appl. Math.* **28** (1975), 201–228. MR 55 #4042 Zbl 0291.31002

Received April 17, 2006.

JUI-TANG RAY CHEN
DEPARTMENT OF MATHEMATICS
NATIONAL TSING HUA UNIVERSITY
HSINCHU
TAIWAN 30013
jtchen@math.nthu.edu.tw

CHIUNG-JUE ANNA SUNG
DEPARTMENT OF MATHEMATICS
NATIONAL TSING HUA UNIVERSITY
HSINCHU
TAIWAN 30013
cjsung@math.nthu.edu.tw
www.math.nthu.edu.tw/∼cjsung

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

TRANSVERSE POISSON STRUCTURES TO ADJOINT ORBITS IN SEMISIMPLE LIE ALGEBRAS

PANTELIS A. DAMIANOU, HERVÉ SABOURIN AND POL VANHAECKE

We study the transverse Poisson structure to adjoint orbits in a complex semisimple Lie algebra. The problem is first reduced to the case of nilpotent orbits. We prove then that in suitably chosen quasihomogeneous coordinates, the quasidegree of the transverse Poisson structure is -2. For *s*ubregular nilpotent orbits, we show that the structure may be computed using a simple determinantal formula that involves the restriction of the Chevalley invariants on the slice. In addition, using results of Brieskorn and Slodowy, the Poisson structure is reduced to a three dimensional Poisson bracket, which is intimately related to the simple rational singularity that corresponds to the subregular orbit.

1. Introduction

The transverse Poisson structure was introduced by A. Weinstein [1983], stating in his famous splitting theorem that every (real smooth or complex holomorphic) Poisson manifold M is, in the neighborhood of each point m, the product of a symplectic manifold and a Poisson manifold of rank 0 at m. The two factors of this product can be geometrically realized as follows. Let S be the symplectic leaf through m, and let N be any submanifold of M containing m such that

$$T_m(M) = T_m(S) \oplus T_m(N).$$

There exists a neighborhood V of m in N, endowed with a Poisson structure, and a neighborhood U of m in S such that, near m, M is isomorphic to the product Poisson manifold $U \times V$. The submanifold N is called a *transverse slice* at m to the symplectic leaf S. The Poisson structure on $V \subset N$ is called the *transverse Poisson structure* to S; up to Poisson isomorphism, it is independent of the point $m \in S$ and the chosen transverse slice N at m: given two points m, $m' \in S$ with

MSC2000: primary 53D17; secondary 17B10, 14J17.
Keywords: transverse Poisson structure, nilpotent orbits, Kleinian singularities.
The authors would like to thank the Cyprus Research Foundation and the Ministère Français des Affaires étrangères for their support.

transverse slices N, N' to S, there exist neighborhoods V of m in N and V' of m' in N' such that (V, m) and (V', m') are Poisson diffeomorphic.

When M is the dual $\mathfrak{g}^*$ of a complex Lie algebra $\mathfrak{g}$ and is equipped with its standard Lie–Poisson structure, we know that the symplectic leaf through $\mu \in \mathfrak{g}^*$ is the coadjoint orbit $\mathrm{G} \cdot \mu$ of the adjoint Lie group G of $\mathfrak{g}$. In this case, a natural transverse slice to $\mathrm{G} \cdot \mu$ is obtained in the following way. We choose any complement $\mathfrak{n}$ to the centralizer $\mathfrak{g}(\mu)$ of μ in $\mathfrak{g}$, and we take N to be the affine subspace $\mu + \mathfrak{n}^\perp$ of $\mathfrak{g}^*$. Since $\mathfrak{g}(\mu)^\perp = \mathrm{ad}^*_{\mathfrak{g}}\,\mu$, we have

$$T_\mu(\mathfrak{g}^*) = T_\mu(\mathrm{G} \cdot \mu) \oplus T_\mu(N),$$

so that N is indeed a transverse slice to $\mathrm{G} \cdot \mu$ at μ. Furthermore, defining on $\mathfrak{n}^\perp$ any system of linear coordinates $(q_1, \dots, q_k)$ and using the explicit formula for Dirac reduction (see formula (4) below), one can write down explicit formulas for the Poisson matrix $\Lambda_N := \left(\left\{q_i, q_j\right\}_N\right)$, $1 \le i, j \le k$ of the transverse Poisson structure, from which it follows easily that the coefficients of Λ_N are actually rational functions in $(q_1, \dots, q_k)$. As a corollary, in the Lie–Poisson case, the transverse Poisson structure is always rational [Saint-Germain 1999]. One immediately wonders, for which cases — more precisely, for which Lie algebras $\mathfrak{g}$, coadjoint orbits, and complements $\mathfrak{n}$ — is the Poisson structure on N polynomial?

Partial answers have been given in the literature for (co)adjoint orbits in a semisimple Lie algebra. P. Damianou [1996] computed explicitly how the transverse Poisson structure to nilpotent orbits of $\mathfrak{gl}_n$ for $n \le 7$ correspond to a particular complement $\mathfrak{n}$; in this case the transverse Poisson structure is polynomial. Cushman and Roberts [2002] proved that there exists for any nilpotent adjoint orbit of a semisimple Lie algebra a special choice of a complement $\mathfrak{n}$ such that the corresponding transverse Poisson structure is polynomial. For the latter case, H. Sabourin [2005] gave a more general class of complements having a polynomial transverse structure, using essentially the machinery of semisimple Lie algebras; he also showed that the choice of complement $\mathfrak{n}$ is relevant for the polynomial character of the transverse Poisson structure by giving an example where the structure is rational for a generic choice of complement.

When the transverse Poisson structure is polynomial, one is tempted to define its degree as the maximal degree of the coefficients $\left\{q_i, q_j\right\}_N$ of its Poisson matrix, as was done in [Damianou 1996] and [Cushman and Roberts 2002], where several conjectures about this degree are formulated. Unfortunately, as shown in [Sabourin 2005], this degree depends strongly on the choice of the complement $\mathfrak{n}$, and hence it is not intrinsically attached to the transverse Poisson structure. We show in Section 3 that the right approach is to use the more general notion of quasidegree; that is, we assign natural quasidegrees $\varpi(q_i)$ to the variables q_i $(i = 1, \dots, k)$ and we show that, in the above mentioned class of complements, the quasidegree of the transverse Poisson structure is always -2, irrespective of the simple Lie algebra, the chosen adjoint orbit, and the chosen transverse slice N! In fact, the weights $\varpi(q_i)$ have a Lie-theoretic origin and are also independent of the particular complement. It follows that $\{q_i, q_j\}_N$ for $1 \le i, j \le k$ is a quasihomogeneous polynomial of quasidegree $\varpi(q_i) + \varpi(q_j) - 2$.

Another result, established in this article, is that the study of the transverse Poisson structure to any adjoint orbit $\mathrm{G}\cdot x$ can be reduced, via the Jordan–Chevalley decomposition of $x \in \mathfrak{g}$, to the case of an adjoint *nilpotent* orbit. Thereby we explain why we are merely interested in the case of nilpotent orbits.

The transverse structure to the regular nilpotent orbit $\mathbb{O}_{reg}$ of $\mathfrak{g}$ is always trivial. So, the next step is to consider the case of the *subregular* nilpotent orbit $\mathbb{O}_{sr}$ of $\mathfrak{g}$. Then $N \cong \mathbb{C}^{\ell+2}$, where ℓ is the rank of $\mathfrak{g}$. The dimension of $\mathbb{O}_{sr}$ is two less than the dimension of the regular orbit, so that the transverse Poisson structure has rank 2. It has ℓ independent polynomial Casimir functions $\chi_1, \dots, \chi_\ell$, where χ_i is the restriction of the i-th Chevalley invariant G_i to the slice N. In this case, the transverse Poisson structure may be obtained by a simple determinantal formula instead of the usual, rather complicated Dirac constraints. That formula is as follows. In linear coordinates $q_1, q_2, \dots, q_{\ell+2}$ on N,

$$\{f, g\}_{\det} := \frac{df \wedge dg \wedge d\chi_1 \wedge \cdots \wedge d\chi_\ell}{dq_1 \wedge dq_2 \wedge \ldots \wedge dq_{\ell+2}} \tag{1}$$

defines a Poisson bracket on N that coincides (up to a nonzero constant) with the transverse Poisson structure on N.

As an application of formula (1), we show in Theorem 5.6 that the Poisson matrix of the transverse Poisson on N takes, in suitable coordinates, the block form

$$\widetilde{\Lambda}_N = \begin{pmatrix} 0 & 0 \\ 0 & \Omega \end{pmatrix}, \quad \text{where} \quad \Omega = \begin{pmatrix} 0 & \dfrac{\partial F}{\partial q_{\ell+2}} & -\dfrac{\partial F}{\partial q_{\ell+1}} \\ -\dfrac{\partial F}{\partial q_{\ell+2}} & 0 & \dfrac{\partial F}{\partial q_\ell} \\ \dfrac{\partial F}{\partial q_{\ell+1}} & -\dfrac{\partial F}{\partial q_\ell} & 0 \end{pmatrix}.$$

The polynomial $F = F(u_1, \dots, u_{\ell-1}, q_\ell, q_{\ell+1}, q_{\ell+2})$ is precisely the one that describes the universal deformation of the (homogeneous or inhomogeneous) simple singularity of the singular surface $N \cap \mathcal{N}$, where $\mathcal{N}$ is the nilpotent cone of $\mathfrak{g}$. The $u_1, \dots, u_{\ell-1}$ are the deformation parameters, which are also Casimirs for the Poisson structure on N. In particular, the restriction of this Poisson structure to

$N \cap \mathcal{N}$ is given by

$$\{x, y\} = \frac{\partial F_0}{\partial z}, \qquad \{y, z\} = \frac{\partial F_0}{\partial x}, \qquad \{z, x\} = \frac{\partial F_0}{\partial y},$$

where $F_0(x, y, z) := F(0, \dots, 0, x, y, z)$ is the polynomial that defines $N \cap \mathcal{N}$ as a surface in $\mathbb{C}^3$. As we will recall in Section 5, Brieskorn [1971] showed that, in the ADE case, the so-called adjoint quotient $G = (G_1, \dots, G_\ell) : \mathfrak{g} \to \mathbb{C}^\ell$ is, when restricted to the slice N, a semiuniversal deformation of the singular surface $N \cap \mathcal{N}$; this result was generalized by Slodowy [1980a] to the other simple Lie algebras. Our Theorem 5.6 adds a Poisson dimension to this result.

The article is organized as follows. In Section 2, we recall a few basic facts concerning transverse Poisson structures, and we show that a general orbit in a semisimple Lie algebra can be reduced to the case of a nilpotent orbit. In Section 3, we recall the notion of quasihomogeneity, and we show that, for a natural class of slices, the transverse Poisson structure is quasihomogeneous of quasidegree -2. In Section 4 and the end of Section 5, we show, in the Lie algebras $\mathfrak{g}_2$, $\mathfrak{so}_8$, and $\mathfrak{sl}_4$, how the transverse Poisson structure can be computed explicitly, and we use these examples to illustrate our results. In Section 5, we prove that, in the case of the subregular orbit, the transverse Poisson structure is given by a determinantal formula; we also show that this Poisson structure is entirely determined by the singular variety of nilpotent elements of the slice.

2. Transverse Poisson structures in semisimple Lie algebras

In this section, we recall the main setup for studying the transverse Poisson structure to a (co)adjoint orbit of a complex semisimple Lie algebra $\mathfrak{g}$, and we show how the general orbit is related to the case of a nilpotent orbit. We use the Killing form $\langle \cdot \mid \cdot \rangle$ of $\mathfrak{g}$ to identify $\mathfrak{g}$ with its dual $\mathfrak{g}^*$. This leads to a Poisson structure on $\mathfrak{g}$ that is given for functions F, G on $\mathfrak{g}$ at $x \in \mathfrak{g}$ by

$$\{F, G\}(x) := \langle x \mid [dF(x), dG(x)] \rangle, \tag{2}$$

where we think of $dF(x)$ and $dG(x)$ as elements of $\mathfrak{g} \cong \mathfrak{g}^* \cong T_x^*\mathfrak{g}$. Since the Killing form is Ad-invariant, the isomorphism $\mathfrak{g} \cong \mathfrak{g}^*$ identifies the adjoint orbits $\mathrm{G} \cdot x$ of G with the coadjoint orbits $\mathrm{G} \cdot \mu$, and so the symplectic leaf of $\{\cdot, \cdot\}$ that passes through x is the adjoint orbit $\mathrm{G} \cdot x$. Also, as a transverse slice at x to $\mathrm{G} \cdot x$, we can take an affine subspace $N := x + \mathfrak{n}^\perp$, where $\mathfrak{n}$ is any complementary subspace to the centralizer $\mathfrak{g}(x) := \{y \in \mathfrak{g} \mid [x, y] = 0\}$ of x in $\mathfrak{g}$ and $\perp$ is the orthogonal complement with respect to the Killing form. To give an explicit formula for the Poisson structure $\{\cdot, \cdot\}_N$ transverse to $\mathrm{G} \cdot x$, let $(Z_1, \dots, Z_k)$ be a basis for $\mathfrak{g}(x)$, and let $(X_1, \dots, X_{2r})$ be a basis for $\mathfrak{n}$, where $2r = \dim(\mathrm{G} \cdot x)$ is the rank of the Poisson structure (2) at x. These bases lead to linear coordinates $q_1, \dots, q_{k+2r}$

on $\mathfrak{g}$, centered at x, defined by $q_i(y) := \langle y - x \mid Z_i\rangle$, for $i = 1, \ldots, k$ and $q_{k+i}(y) := \langle y - x \mid X_i\rangle$, for $i = 1, \ldots, 2r$. Since $dq_i(y) = Z_i$ for $i = 1, \ldots, k$ and $dq_{k+i}(y) = X_i$ for $i = 1, \ldots, 2r$, it follows from (2) that the Poisson matrix of $\{\cdot, \cdot\}$ at $y \in \mathfrak{g}$ is given by

$$\left(\left\{q_i, q_j\right\}(y)\right)_{1\le i,j\le k+2r} = \begin{pmatrix} A(y) & B(y) \\ -B(y)^\top & C(y) \end{pmatrix}, \tag{3}$$

where

$$\begin{aligned} A_{i,j}(y) &= \left\langle y \mid [Z_i, Z_j]\right\rangle, && \text{for } 1 \le i, j \le k; \\ B_{i,m}(y) &= \langle y \mid [Z_i, X_m]\rangle, && \text{for } 1 \le i \le k, \quad 1 \le m \le 2r; \\ C_{l,m}(y) &= \langle y \mid [X_l, X_m]\rangle, && \text{for } 1 \le l, m \le 2r. \end{aligned}$$

It is easy to see that the skew-symmetric matrix $C(x)$ is invertible, and so $C(y)$ is invertible for y in a neighborhood of x in $\mathfrak{g}$, and hence for y in a neighborhood V of x in N. By Dirac reduction, the Poisson matrix of $\{\cdot, \cdot\}_N$ at $n \in V$ in the coordinates $q_1, \ldots, q_k$ (restricted to V), is given by

$$\Lambda_N(n) = A(n) + B(n)C(n)^{-1}B(n)^\top. \tag{4}$$

According to the Jordan–Chevalley decomposition theorem, we can write $x = s + e$, where s is semisimple, e is nilpotent, and $[s, e] = 0$. Moreover, the respective centralizers of x, s and e are related as follows:

$$\mathfrak{g}(x) = \mathfrak{g}(s) \cap \mathfrak{g}(e). \tag{5}$$

This leads to a natural class of complements $\mathfrak{n}$ to $\mathfrak{g}(x)$. Since the restriction of $\langle \cdot \mid \cdot \rangle$ to $\mathfrak{g}(s)$ is nondegenerate [Dixmier 1996, Prop. 1.7.7.], we have a vector space decomposition of $\mathfrak{g}$ as

$$\mathfrak{g} = \mathfrak{g}(s) \oplus \mathfrak{n}_s,$$

where $\mathfrak{n}_s = \mathfrak{g}(s)^\perp$. Notice that $\mathfrak{n}_s$ is $\mathfrak{g}(s)$-invariant, that is, $[\mathfrak{g}(s), \mathfrak{n}_s] \subset \mathfrak{n}_s$, since

$$\langle \mathfrak{g}(s) \mid [\mathfrak{g}(s), \mathfrak{n}_s]\rangle = \langle [\mathfrak{g}(s), \mathfrak{g}(s)] \mid \mathfrak{n}_s\rangle \subset \langle \mathfrak{g}(s) \mid \mathfrak{n}_s\rangle = \{0\}.$$

Choosing any complement $\mathfrak{n}_e$ of $\mathfrak{g}(x)$ in $\mathfrak{g}(s)$, we get the following decomposition of $\mathfrak{g}$:

$$\mathfrak{g} = \mathfrak{g}(x) \oplus \mathfrak{n}_e \oplus \mathfrak{n}_s.$$

We take then $\mathfrak{n} := \mathfrak{n}_e \oplus \mathfrak{n}_s$, and we denote $N_x := x + \mathfrak{n}^\perp$. It follows that, if $n \in N_x$ such that $n \in \mathfrak{g}(s)$, then $\langle n \mid [\mathfrak{g}(s), \mathfrak{n}_s]\rangle \subset \langle \mathfrak{g}(s) \mid \mathfrak{n}_s\rangle = \{0\}$. In particular,

$$\langle n \mid [\mathfrak{g}(x), \mathfrak{n}_s]\rangle = \{0\} \quad \text{and} \quad \langle n \mid [\mathfrak{n}_e, \mathfrak{n}_s]\rangle = \{0\}. \tag{6}$$

Let us assume that the basis vectors $X_1, \ldots, X_{2r}$ of $\mathfrak{n}$ have been chosen such that $X_1, \ldots, X_{2p} \in \mathfrak{n}_e$ and $X_{2p+1}, \ldots, X_{2r} \in \mathfrak{n}_s$. Then the formulas (6) imply that the

Poisson matrix (3) takes at $n \in N_x$ the form

$$\Lambda(n) = \begin{pmatrix} A(n) & B_e(n) & 0 \\ -B_e(n)^\top & C_e(n) & 0 \\ 0 & 0 & C_s(n) \end{pmatrix},$$

where

$$\begin{aligned} A_{i,j}(n) &= \big\langle n \mid [Z_i, Z_j] \big\rangle, && \text{for } 1 \le i, j \le k; \\ B_{e;i,m}(n) &= \langle n \mid [Z_i, X_m] \rangle, && \text{for } 1 \le i \le k,\ 1 \le m \le 2p; \\ C_{e;l,m}(n) &= \langle n \mid [X_l, X_m] \rangle, && \text{for } 1 \le l,\ m \le 2p; \\ C_{s;l,m}(n) &= \langle n \mid [X_l, X_m] \rangle, && \text{for } 2p+1 < l,\ m \le 2r. \end{aligned}$$

It follows from (4) that the Poisson matrix of the transverse Poisson structure on N_x is given by

$$\Lambda_{N_x}(n) = A(n) + B_e(n) C_e(n)^{-1} B_e(n)^\top. \tag{7}$$

Let us now restrict our attention to the Lie algebra $\mathfrak{g}(s)$, which, being reductive, decomposes as

$$\mathfrak{g}(s) = \mathfrak{z}(s) \oplus \mathfrak{g}_{ss}(s),$$

where $\mathfrak{z}(s)$ is the center of $\mathfrak{g}(s)$ and $\mathfrak{g}_{ss}(s) = [\mathfrak{g}(s), \mathfrak{g}(s)]$ is the semisimple part of $\mathfrak{g}(s)$. At the group level we have a similar decomposition of $\mathrm{G}(s)$, the centralizer of s in G whose Lie algebra is $\mathfrak{g}(s)$, namely,

$$\mathrm{G}(s) = \boldsymbol{Z}(s)\mathrm{G}_{ss}(s),$$

where $\boldsymbol{Z}(s)$ is a central subgroup of $\mathrm{G}(s)$ and $\mathrm{G}_{ss}(s)$ is the semisimple part of $\mathrm{G}(s)$ with Lie algebra $\mathfrak{g}_{ss}(s)$. Since $e \in \mathfrak{g}(s)$, we can consider $\mathrm{G}(s) \cdot e$ as an adjoint orbit of the reductive Lie algebra $\mathfrak{g}(s)$. We may think of it as an adjoint orbit of a *semisimple* Lie algebra, since $\mathrm{G}(s) \cdot e = \mathrm{G}_{ss}(s) \cdot e$; similarly we may think of a transverse slice to the adjoint orbit $\mathrm{G}(s) \cdot e$ as a transverse slice to $\mathrm{G}_{ss}(s) \cdot e$ up to a summand with trivial Lie bracket. Denoting by $\perp_s$ the $\langle \cdot \mid \cdot \rangle$ orthogonal complement restricted to $\mathfrak{g}(s)$, we have that $N := e + \mathfrak{n}_e^{\perp_s}$ is a transverse slice to $\mathrm{G}(s) \cdot e$, since

$$\mathfrak{g}(s) = \mathfrak{g}(x) \oplus \mathfrak{n}_e = \mathfrak{z}(s) \oplus \mathfrak{g}_{ss}(s)(e) \oplus \mathfrak{n}_e.$$

We have used that $\mathfrak{g}(x) = \mathfrak{g}(s)(e)$ is the centralizer of e in $\mathfrak{g}(s)$, which follows from (5). In the chosen bases $(Z_1, \ldots, Z_k)$ of $\mathfrak{g}(x)$ and $(X_1, \ldots, X_{2p})$ of $\mathfrak{n}_e$, the Poisson matrix at $n \in N$ takes the form

$$\begin{pmatrix} A(n) & B_e(n) \\ -B_e(n)^\top & C_e(n) \end{pmatrix},$$

which leads by Dirac reduction to the transverse Poisson structure Λ_N on N:

$$\Lambda_N(n) = A(n) + B_e(n)C_e(n)^{-1}B_e(n)^\top,$$

where $n \in N$. This yields formally the same formula as (7), except that it is evaluated at points n of N rather than at points of N_x. However, since $\mathfrak{n}_e^{\perp_s} = \mathfrak{g}(s) \cap \mathfrak{n}_e^\perp = \mathfrak{n}_s^\perp \cap \mathfrak{n}_e^\perp = (\mathfrak{n}_s + \mathfrak{n}_e)^\perp = \mathfrak{n}^\perp$, the affine subspaces N_x and N only differ by a translation, $N_x = s + e + \mathfrak{n}^\perp = s + N$. Thus they, and their Poisson matrices with respect to the coordinates $q_1, \dots, q_k$, can be identified, leading to:

Proposition 2.1. *Let $x \in \mathfrak{g}$ be any element, $\mathrm{G}\cdot x$ its adjoint orbit, and $x = s + e$ its Jordan–Chevalley decomposition. Given any complement $\mathfrak{n}_e$ of $\mathfrak{g}(x)$ in $\mathfrak{g}(s)$ and putting $\mathfrak{n} := \mathfrak{n}_s \oplus \mathfrak{n}_e$, where $\mathfrak{n}_s = \mathfrak{g}(s)^\perp$, the parallel affine spaces $N_x := x + \mathfrak{n}^\perp$ and $N := e + \mathfrak{n}^\perp$ are respectively transverse slices to the adjoint orbit $\mathrm{G}\cdot x$ in $\mathfrak{g}$ and to the nilpotent orbit $\mathrm{G}(s)\cdot e$ in $\mathfrak{g}(s)$. The Poisson structure on both transverse slices has the same Poisson matrix, namely that of* (7), *in the same affine coordinates restricted to the corresponding transverse slice.*

In short, the transverse Poisson structure to any adjoint orbit $\mathrm{G}\cdot x$ of a semisimple (or reductive) Lie algebra $\mathfrak{g}$ is essentially determined by the transverse Poisson structure of the underlying nilpotent orbit $\mathrm{G}(s)\cdot e$ defined by the Jordan–Chevalley decomposition $x = s + e$. A refinement of this proposition will be given in Corollary 3.5.

3. The polynomial and the quasihomogeneous character of the tranverse Poisson structure

In this section we show that, for a natural class of transverse slices to a nilpotent orbit $\mathbb{O}$ which we equip with an adapted set of linear coordinates centered at a nilpotent element $e \in \mathbb{O}$, the transverse Poisson structure is quasihomogeneous (of quasidegree -2) in the following sense.

Definition 3.1. Let $\nu = (\nu_1, \dots, \nu_d)$ be nonnegative integers. A polynomial P in $\mathbb{C}[x_1, \dots, x_d]$ is said to be *quasihomogeneous* (relative to ν) if, for some integer κ,

$$P(t^{\nu_1}x_1, \dots, t^{\nu_d}x_d) = t^\kappa P(x_1, \dots, x_d) \quad \text{for all } t \in \mathbb{C},$$

and κ is then called the *quasidegree* (relative to ν) of P, denoted $\varpi(P)$. Similarly, a polynomial Poisson structure $\{\cdot\,,\cdot\}$ on $\mathbb{C}[x_1, \dots, x_d]$ is said to be *quasihomogeneous* (relative to ν) if there exists $\kappa \in \mathbb{Z}$ such that, for any quasihomogeneous polynomials F and G, their Poisson bracket $\{F, G\}$ is quasihomogeneous of degree

$$\varpi(\{F, G\}) = \varpi(F) + \varpi(G) + \kappa;$$

equivalently, for any i, j the polynomial $\left\{x_i, x_j\right\}$ is quasihomogeneous of quasidegree $\nu_i + \nu_j + \kappa$. Then κ is called the *quasidegree* of $\{\cdot\,,\cdot\}$.

We first show that, given $\mathbb{O}$, we can choose a system of linear coordinates on $\mathfrak{g}$, centered at some nilpotent element $e \in \mathbb{O}$, such that the Lie–Poisson structure on $\mathfrak{g}$ is quasihomogeneous relative to some vector ν that has a natural Lie-theoretic interpretation. To describe how this happens, we need to recall some facts from the theory of semisimple Lie algebras, which will be used throughout this paper. First, one chooses a Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g}$, with corresponding root system $\Delta(\mathfrak{h})$, from which a basis $\Pi(\mathfrak{h})$ of simple roots is selected. The *rank* of $\mathfrak{g}$, which is the dimension of $\mathfrak{h}$, is denoted by ℓ. According to the Jacobson–Morosov–Kostant correspondence (see [Tauvel and Yu 2005, paragraphs 32.1 and 32.4]), there is a canonical triple $(h, e, f) \in \mathfrak{g}$ associated with $\mathbb{O}$ and completely determined, up to conjugation by $G(h)$, by the following properties:

- (h, e, f) is a $\mathfrak{sl}_2$-triple, that is, $[h, e] = 2e$, $[h, f] = -2f$, and $[e, f] = h$;
- h is the characteristic of $\mathbb{O}$, that is, $h \in \mathfrak{h}$ and $\alpha(h) \in \{0, 1, 2\}$ for any simple root $\alpha \in \Pi(\mathfrak{h})$.
- $\mathbb{O} = \mathrm{G} \cdot e$.

The triple (h, e, f) leads to two decompositions of $\mathfrak{g}$.

First, $\mathfrak{g}$ decomposes into eigenspaces relative to ad_h. Since each eigenvalue is an integer, we have

$$\mathfrak{g} = \bigoplus_{i \in \mathbb{Z}} \mathfrak{g}(i),$$

where $\mathfrak{g}(i)$ is the eigenspace of ad_h with eigenvalue i. For example, $e \in \mathfrak{g}(2)$ and $f \in \mathfrak{g}(-2)$.

Second, let $\mathfrak{s}$ be the Lie subalgebra of $\mathfrak{g}$ isomorphic to $\mathfrak{sl}_2$ that is generated by h, e and f. The Lie algebra $\mathfrak{g}$ is an $\mathfrak{s}$-module, hence it decomposes as

$$\mathfrak{g} = \bigoplus_{j=1}^{k} V_{n_j},$$

where each V_{n_j} is a simple $\mathfrak{s}$-module, with $n_j + 1 = \dim V_{n_j}$ and ad_h-weights $n_j, n_j - 2, n_j - 4, \dots, -n_j$. Moreover, $k = \dim \mathfrak{g}(e)$, since the centralizer $\mathfrak{g}(e)$ is generated by the highest weight vectors of each V_{n_j}. It follows that

$$\sum_{j=1}^{k} n_j = \dim \mathfrak{g} - k = \dim(\mathrm{G} \cdot e) = 2r. \tag{8}$$

We center at e a system of linear coordinates on $\mathfrak{g}$ by using the action of Slodowy [1980b]: First, he considers the one-parameter subgroup of G,

$$\begin{array}{rccl} \lambda : & \mathbb{C}^* & \to & \mathrm{G} \\ & t & \mapsto & \exp(\lambda_t h), \end{array}$$

where λ_t is a complex number such that $e^{-\lambda_t} = t$. The restriction of Ad to this subgroup leaves every eigenspace $\mathfrak{g}(i)$ invariant and acts for each t as a homothecy with ratio t^{-i} on $\mathfrak{g}(i)$:

$$\mathrm{Ad}_{\lambda(t)}\, x = t^{-i} x \quad \text{for all } x \in \mathfrak{g}(i). \tag{9}$$

Since $e \in \mathfrak{g}(2)$, the action ρ of $\mathbb{C}^*$ on $\mathfrak{g}$ — defined for $t \in \mathbb{C}^*$ and for $y \in \mathfrak{g}$ by $\rho_t \cdot y := t^2\,\mathrm{Ad}_{\lambda(t)}\, y$ — fixes e. We refer to ρ as *Slodowy's action.* To see how it leads to quasihomogeneous coordinates, let us define for $x \in \mathfrak{g}$ the function $\mathcal{F}_x(y) := \langle y - e \mid x\rangle$ for $y \in \mathfrak{g}$. Then (9) and the Ad-invariance of the Killing form imply that if $x \in \mathfrak{g}(i)$ then

$$\begin{aligned}\left(\rho_t^*\mathcal{F}_x\right)(y) &= \left\langle \rho_{t^{-1}} \cdot y - e \mid x\right\rangle = t^{-2}\left\langle \mathrm{Ad}_{\lambda(t^{-1})}(y-e) \mid x\right\rangle \\ &= t^{-2}\left\langle y - e \mid \mathrm{Ad}_{\lambda(t)}\, x\right\rangle = t^{-2}\left\langle y - e \mid t^{-i}x\right\rangle = t^{-i-2}\mathcal{F}_x(y).\end{aligned}$$

It follows that the quasidegree $\varpi(\mathcal{F}_x)$ of $\mathcal{F}_x$ is $i+2$ for $x \in \mathfrak{g}(i)$. According to (2), one has, for any $x,\ y,\ z \in \mathfrak{g}$,

$$\left\{\mathcal{F}_x, \mathcal{F}_y\right\}(z) = \langle z \mid [x, y]\rangle = \mathcal{F}_{[x,y]}(z) + \langle e \mid [x, y]\rangle\,. \tag{10}$$

If $x \in \mathfrak{g}(i)$ and $y \in \mathfrak{g}(j)$ with $i + j \neq -2$, then $\langle e \mid [x, y]\rangle = 0$ and so

$$\begin{aligned}\varpi\left(\left\{\mathcal{F}_x, \mathcal{F}_y\right\}\right) - \varpi(\mathcal{F}_x) - \varpi(\mathcal{F}_y) &= \varpi(\mathcal{F}_{[x,y]}) - \varpi(\mathcal{F}_x) - \varpi(\mathcal{F}_y) \\ &= i + j + 2 - (i+2) - (j+2) = -2.\end{aligned}$$

This result extends to the case $i + j = -2$, since then $\varpi(\mathcal{F}_{[x,y]}) = i + j + 2 = 0$, which is the quasidegree of the constant function $\langle e \mid [x, y]\rangle$. This proves:

Proposition 3.2. *Let $\mathfrak{g}$ be a semisimple Lie algebra identified with its dual using its Killing form. Let $\mathcal{O}$ be a nilpotent adjoint orbit of $\mathfrak{g}$ with canonical triple (h, e, f). Let $x_1, \dots, x_d$ be any basis in $\mathfrak{g}$, where each x_k belongs to some eigenspace $\mathfrak{g}(i_k)$ of* ad_h, *and let $\mathcal{F}_k$ be the dual coordinates on $\mathfrak{g}$ centered at e as $\mathcal{F}_k(y) := \langle y - e \mid x_k\rangle$. Then the Lie–Poisson structure $\{\cdot, \cdot\}$ on $\mathfrak{g}$ is quasihomogeneous of degree -2 with respect to $(\varpi(\mathcal{F}_1), \dots, \varpi(\mathcal{F}_d)) = (i_1 + 2, \dots, i_d + 2)$.* □

We now wish to show that, upon picking a suitable transverse slice N to $\mathcal{O}$ at e, the transverse Poisson structure on N is also quasihomogeneous (of degree -2). Following [Sabourin 2005], we consider the set $\mathcal{N}_h$ of all subspaces $\mathfrak{n}$ of $\mathfrak{g}$ that are complementary to $\mathfrak{g}(e)$ in $\mathfrak{g}$ and are ad_h-invariant. For $\mathfrak{n} \in \mathcal{N}_h$ we let $N := e + \mathfrak{n}^\perp$, which is a transverse slice to $\mathrm{G}\cdot e$. The ad_h-invariance of $\mathfrak{n}$ implies on the one hand that ρ leaves N invariant: if $y \in e + \mathfrak{n}^\perp$ then

$$0 = \left\langle y - e \mid \mathrm{Ad}_{\lambda(t^{-1})}\, \mathfrak{n}\right\rangle = \left\langle \mathrm{Ad}_{\lambda(t)}(y-e) \mid \mathfrak{n}\right\rangle = t^{-2}\,\langle \rho_t \cdot y - e \mid \mathfrak{n}\rangle\,,$$

so that indeed $\rho_t \cdot y \in e + \mathfrak{n}^\perp$. On the other hand, it implies that $\mathfrak{n}$ admits a basis consisting of eigenvectors of $\mathfrak{h}$. Thus we can adapt the above basis $x_1, \dots, x_d$ to $\mathfrak{n}$.

We can choose a basis $(Z_1, \dots, Z_k)$ for $\mathfrak{g}(e)$ and a basis $(X_1, \dots, X_{2r})$ for $\mathfrak{n}$ so that:

- each Z_i for $1 \le i \le k$ is a highest weight vector of weight n_i;
- each X_i for $1 \le i \le 2r$ is a weight vector of weight ν_i.

The linear coordinates (centered at e) $\mathcal{F}_{Z_1}, \dots, \mathcal{F}_{Z_k}$, when restricted to N, will be denoted $q_1, \dots, q_k$. By the above, their quasidegrees are defined as $\varpi(q_i) := n_i + 2$. That the transverse Poisson structure is polynomial in these coordinates was first shown in [Sabourin 2005, Thm 2.3]. We now refine this statement.

Proposition 3.3. *In the notation of Proposition 3.2, the transverse Poisson structure on $N := e + \mathfrak{n}^\perp$, where $\mathfrak{n} \in \mathcal{N}$, is a polynomial Poisson structure that is quasihomogeneous of degree -2 with respect to the quasidegrees $n_1 + 2, \dots, n_k + 2$, where $n_1, \dots, n_k$ denote the highest weights of $\mathfrak{g}$ as an $\mathfrak{s}$-module.*

Proof. According to (4), we need to show that for any $1 \le i, j \le k$ the functions A_{ij} and $(BC^{-1}B^\top)_{ij}$ are quasihomogeneous of degree $\varpi(q_i) + \varpi(q_j) - 2 = n_i + n_j + 2$. For A_{ij} this is clear, since A is part of the Poisson matrix of the Lie–Poisson structure on $\mathfrak{g}$, which we know is quasihomogeneous of degree -2. Similarly, we have $\varpi(B_{ip}) = n_i + \nu_p + 2$. Since

$$\varpi(B_{ip} C_{ps}^{-1} B_{js}) = n_i + n_j + \nu_p + \nu_s + 4 + \varpi(C_{ps}^{-1}),$$

we must show that

$$\varpi(C_{ps}^{-1}) = -\nu_p - \nu_s - 2. \tag{11}$$

This follows from $\sum_{i=1}^{2r} (\nu_i + 1) = 0$, which is itself a consequence of (8). Indeed, consider a term of the form $C'_{ij} = C_{i_1 j_1} \dots C_{i_{2r-1} j_{2r-1}}$, where

$$\{i_1, i_2, \dots, i_{2r-1}\} = \{1, 2, \dots, 2r\} \setminus \{s\},$$
$$\{j_1, i_2, \dots, j_{2r-1}\} = \{1, 2, \dots, 2r\} \setminus \{p\}.$$

Then

$$\varpi(C'_{ij}) = \sum_{k=1}^{2r-1} (\nu_{i_k} + \nu_{j_k} + 2) = -\nu_s - \nu_p - 2,$$

A typical term of C_{ps}^{-1} is of the form $C'_{ij}/\Delta(C)$, where $\Delta(C)$ is the determinant of C. As C is of quasidegree zero, $\Delta(C)$ is constant by the previous argument. This observation was made in [Sabourin 2005, Theorem 2.3]. This gives us (11). □

Remark 3.4. Our referee pointed out that the quasihomogeneity of the transverse Poisson structure is implicit in [Gan and Ginzburg 2002] and [Premet 2002] for the special transversal $\mathfrak{n} = \operatorname{Ker} \operatorname{ad}_f$. Using the eigenspaces of ad_h, these authors consider a filtration on the universal enveloping algebra $\mathcal{U}\mathfrak{g}$ of $\mathfrak{g}$, which yields a grading on the transversal Poisson algebra for this $\mathfrak{n}$. With the quasidegrees that we use, the filtration's graded algebra is $\mathrm{Sym}(\mathfrak{g}^*)$.

Let us consider now any adjoint orbit $\mathrm{G}\cdot x$ and $x = s + e$, the Jordan–Chevalley decomposition of x. We already considered this case in Proposition 2.1. A well-known result [Tauvel and Yu 2005, par. 32.1.7.] says that there exists an $\mathfrak{sl}_2$-triple (h, e, f) such that $[s, h] = [s, f] = 0$. Consequently, (h, e, f) is an $\mathfrak{sl}_2$-triple of the reductive Lie algebra $\mathfrak{g}(s)$, and we can also suppose that, up to conjugation by elements of $\mathrm{G}(s)$, h is the characteristic of $\mathrm{G}(s)\cdot e$. Let $\mathcal{N}_{s,h}$ be the set of all complementary subspaces to $\mathfrak{g}(x)$ in $\mathfrak{g}(s)$ that are ad_h-invariant. Then, by applying Proposition 3.3, we get:

Corollary 3.5. *As in Proposition 2.1, let* $\mathfrak{n}_s = \mathfrak{g}(s)^{\perp}$, $\mathfrak{n}_e \in \mathcal{N}_{s,h}$ *and* $\mathfrak{n} = \mathfrak{n}_s \oplus \mathfrak{n}_e$. *Let* $N_x := x + \mathfrak{n}^{\perp}$, *which is a transverse slice to* $\mathrm{G}\cdot x$. *Then the transverse Poisson structure on* N_x *is polynomial and is quasihomogeneous of quasidegree* -2.

From now on, a transverse Poisson structure given by Proposition 3.3 will be called an *adjoint transverse Poisson structure* or ATP-structure.

4. Examples

We want to show in two examples how to compute the ATP-structure. In the first example, we consider the subregular orbit of $\mathfrak{g}_2$, and we compute it without choosing a representation of $\mathfrak{g}_2$. In the second example, the subregular orbit of $\mathfrak{so}_8$, we use a concrete representation rather than referring to tables of the Lie brackets in a Chevalley basis. These two examples will also serve later to illustrate the results we will prove on the nature of the ATP-structure. Both examples correspond to subregular orbits and lead to two of the simplest nontrivial ATP-structures in the following sense. If $\mathbb{O}$ is an adjoint orbit in $\mathfrak{g}$, then the ATP-structure to $\mathbb{O}$ has rank $\dim\mathfrak{g} - \ell - \dim\mathbb{O}$ at a generic point of any slice transverse to $\mathbb{O}$, since the Lie–Poisson structure on $\mathfrak{g}$ has rank[1] $\dim\mathfrak{g} - \ell$ at a generic point of $\mathfrak{g}$. For the regular nilpotent orbit $\mathbb{O}_{reg}$, the ATP-structure is trivial because $\dim\mathbb{O}_{reg} = \dim\mathfrak{g} - \ell$. So, the first interesting nilpotent orbit to consider is the subregular orbit, denoted by $\mathbb{O}_{sr}$. We recall two well-known facts [Collingwood and McGovern 1993]:

(1) the subregular orbit $\mathbb{O}_{sr}$ is the unique nilpotent orbit that is open and dense in the complement of $\mathbb{O}_{reg}$ in the nilpotent cone;

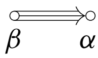
(2) $\dim\mathbb{O}_{sr} = \dim\mathfrak{g} - \ell - 2$.

It follows that the ATP-structure of the subregular orbit had dimension $\ell + 2$ and generic rank 2. In both of the following examples, we give the characteristic triplet

[1] Recall that ℓ denotes the rank of $\mathfrak{g}$.

(h, e, f) that corresponds to the orbit; we derive from it a basis of the ad_h-weight spaces, which leads to basis vectors Z_i of $\mathfrak{g}(e)$ and X_j of an ad_h-invariant complement to $\mathfrak{g}(e)$ in $\mathfrak{g}$. The Lie brackets of these elements then lead to the matrices A, B and C in (3), which, by Dirac's formula (4), yields the matrix Λ_N of the transverse Poisson structure.

The subregular orbit of type G_2. We first consider the case of the subregular orbit of the Lie algebra $\mathfrak{g} := \mathfrak{g}_2$. Denoting the basis of simple roots by $\Pi = \{\alpha, \beta\}$, where β is the longer root, its Dynkin diagram is given by



and it has the positive roots

$$\Delta_+ = \{\alpha,\ \beta,\ \alpha+\beta,\ 2\alpha+\beta,\ 3\alpha+\beta,\ 3\alpha+2\beta\}.$$

The vectors in the Chevalley basis[2] of $\mathfrak{g}$ are denoted by H_α, H_β for the Cartan subalgebra, X_γ for the six positive roots $\gamma \in \Delta_+$, and Y_γ for the six negative roots $-\gamma$, where $\gamma \in \Delta_+$. According to [Collingwood and McGovern 1993, Chapter 8.4], the characteristic h of the subregular orbit $\mathbb{O}_{sr}$ is given by the sequence of weights $(0, 2)$, which means that $\langle \alpha, h\rangle = 0$ and $\langle \beta, h\rangle = 2$ and yields $h = 2H_\alpha + 4H_\beta$. The decomposition of $\mathfrak{g}$ into ad_h-weight spaces $\mathfrak{g}(i)$ consists of five subspaces:

$$\begin{aligned}
\mathfrak{g}(4) &= \langle X_{3\alpha+2\beta}\rangle,\\
\mathfrak{g}(2) &= \langle X_\beta,\ X_{\alpha+\beta},\ X_{2\alpha+\beta},\ X_{3\alpha+\beta}\rangle,\\
\mathfrak{g}(0) &= \langle H_\alpha,\ H_\beta,\ X_\alpha,\ Y_\alpha\rangle,\\
\mathfrak{g}(-2) &= \langle Y_\beta,\ Y_{\alpha+\beta},\ Y_{2\alpha+\beta},\ Y_{3\alpha+\beta}\rangle,\\
\mathfrak{g}(-4) &= \langle Y_{3\alpha+2\beta}\rangle.
\end{aligned}\tag{12}$$

Taking for e and f an arbitrary linear combination of the above basis elements of $\mathfrak{g}(2)$ and $\mathfrak{g}(-2)$, respectively, and using $[e, f] = h$, one easily finds that the $\mathfrak{sl}_2$-triple corresponding to $\mathbb{O}_{sr}$ is

$$e = X_\beta + X_{3\alpha+\beta},\ \ h = 2H_\alpha + 4H_\beta,\ \ f = 2Y_\beta + 2Y_{3\alpha+\beta}.$$

Picking the vectors in the positive subspaces $\mathfrak{g}(i)$ that commute with e leads to basis vectors of $\mathfrak{g}(e)$:

$$\begin{aligned}
Z_1 &= X_\beta + X_{3\alpha+\beta}, & Z_2 &= X_{2\alpha+\beta},\\
Z_3 &= X_{\alpha+\beta}, & Z_4 &= X_{3\alpha+2\beta}.
\end{aligned}\tag{13}$$

[2]The Chevalley basis that we use is explicitly described in [Tauvel 1998, Chapter VII.4].

We obtain an ad_h-invariant complementary subspace $\mathfrak{n}$ of $\mathfrak{g}(e)$ by completing these vectors with additional vectors taken from the bases (12) of the subspaces $\mathfrak{g}(i)$. Our choice of basis vectors for $\mathfrak{n}$, ordered by weight, is

$$\begin{aligned} X_1 &= X_\beta, & X_6 &= Y_\beta, \\ X_2 &= X_\alpha, & X_7 &= Y_{\alpha+\beta}, \\ X_3 &= H_\alpha, & X_8 &= Y_{2\alpha+\beta}, \\ X_4 &= H_\beta, & X_9 &= Y_{3\alpha+\beta}, \\ X_5 &= Y_\alpha, & X_{10} &= Y_{3\alpha+2\beta}. \end{aligned}$$

The Lie brackets of these basis vectors for $\mathfrak{g}$, which are listed in [Tauvel 1998, Chapter VII.4], yield the Poisson matrix $((A, B), (-B^\top, C))$ of the Lie–Poisson structure on $\mathfrak{g}$ in the coordinates $\mathscr{F}_{Z_1}, \dots, \mathscr{F}_{Z_4}, \mathscr{F}_{X_1}, \dots, \mathscr{F}_{X_{10}}$ on $\mathfrak{g}$, as

$$A_{ij} = \left\{\mathscr{F}_{Z_i}, \mathscr{F}_{Z_j}\right\} = \mathscr{F}_{[Z_i, Z_j]} + \left\langle e \mid [Z_i, Z_j]\right\rangle$$

(see (10)), and similarly for the other elements of the Poisson matrix. We give the restriction of the matrices A, B and C to the transverse slice $N := e + \mathfrak{n}^\perp$ only, which amounts to keeping in the Lie brackets only the vectors $Z_1, \dots, Z_4$, as $\mathscr{F}_X(n) = \langle e - n \mid X\rangle = 0$ for $X \in \mathfrak{n}$ and $n \in N = e + \mathfrak{n}^\perp$. In the coordinates $q_1, \dots, q_4$ on N, where q_i is the restriction of $\mathscr{F}_{Z_i}$ to N, we get

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -3q_4 & 0 \\ 0 & 3q_4 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} 0 & 0 & 0 & -q_4 & 0 & q_1 & -q_2 & q_3 & 0 & 0 \\ 0 & 3q_1 & -q_2 & 0 & 2q_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2q_2 & q_3 & -q_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ q_4 & q_3 & -3q_1 & q_1 & q_2 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$C = \frac{1}{3}\begin{pmatrix} 0 & 3q_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ -3q_3 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & -3 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Substituted in (4), this yields the Poisson matrix for the ATP-structure:

$$\Lambda_N = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -3q_4 & 2q_1q_2 - 2q_3^2 \\ 0 & 3q_4 & 0 & 2q_2^2 - 2q_1q_3 \\ 0 & -2q_1q_2 + 2q_3^2 & -2q_2^2 + 2q_1q_3 & 0 \end{pmatrix}. \tag{14}$$

It follows from (12) and (13) that the quasidegree of q_1, q_2 and q_3 is 4, while the quasidegree of q_4 is 6. One easily reads off from (14) that, with respect to these quasidegrees, the ATP-structure is quasihomogeneous of quasidegree -2.

The subregular orbit of type D_4. We now take $\mathfrak{g} = \mathfrak{so}_8$ and we realize $\mathfrak{g}$ as the following set of matrices:

$$\left\{ \begin{pmatrix} Z_1 & Z_2 \\ Z_3 & -Z_1^\top \end{pmatrix} \mid Z_i \in \mathrm{Mat}_4(\mathbb{C}),\ \text{with } Z_2, Z_3 \text{ skew-symmetric} \right\}.$$

Let $\mathfrak{h}$ denote the Cartan subalgebra of $\mathfrak{g}$ consisting of all diagonal matrices in $\mathfrak{g}$. Clearly, $\mathfrak{h}$ is spanned by the four matrices $H_i := E_{i,i} - E_{4+i,4+i}$, $1 \le i \le 4$. Define for $i = 1, \dots, 4$ the linear map $e_i \in \mathfrak{h}^*$ by

$$e_i\left(\sum a_k H_k\right) = a_i.$$

Then the root system of $\mathfrak{g}$ is

$$\Delta := \{\pm e_i \pm e_j \mid 1 \le i, j \le 4, i \ne j\},$$

and a basis of simple roots is $\Pi := \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$, where

$$\alpha_1 = e_1 - e_2,\ \alpha_2 = e_2 - e_3,\ \alpha_3 = e_3 - e_4,\ \alpha_4 = e_3 + e_4.$$

It leads to the following Chevalley basis of $\mathfrak{g}$:

$$\begin{aligned} X_{e_i - e_j} &= E_{i,j} - E_{4+j,4+i}, & \\ X_{e_i + e_j} &= E_{i,4+j} - E_{j,4+i}, & i < j, \\ X_{-e_i - e_j} &= -E_{4+i,j} + E_{4+j,i}, & i < j, \\ H_{e_i - e_j} &= H_i - H_j, & \\ H_{e_i + e_j} &= H_i + H_j. & \end{aligned}$$

According to [Collingwood and McGovern 1993, Chapter 5.4], the characteristic h of the subregular orbit is given by the sequence of weights $(2, 0, 2, 2)$. It follows that

$$h = 4H_{\alpha_1} + 6H_{\alpha_2} + 4H_{\alpha_3} + 4H_{\alpha_4}.$$

The positive ad_h-weight spaces are

$$\begin{aligned}
\mathfrak{g}(0) &= \mathfrak{h} \oplus \langle X_{\alpha_2}, X_{-\alpha_2} \rangle, \\
\mathfrak{g}(2) &= \langle X_{\alpha_1}, X_{\alpha_3}, X_{\alpha_4}, X_{\alpha_1+\alpha_2}, X_{\alpha_2+\alpha_3}, X_{\alpha_2+\alpha_4} \rangle, \\
\mathfrak{g}(4) &= \langle X_{\alpha_1+\alpha_2+\alpha_3}, X_{\alpha_2+\alpha_3+\alpha_4}, X_{\alpha_1+\alpha_2+\alpha_4} \rangle, \\
\mathfrak{g}(6) &= \langle X_{\alpha_1+\alpha_2+\alpha_3+\alpha_4}, X_{\alpha_1+2\alpha_2+\alpha_3+\alpha_4} \rangle.
\end{aligned} \tag{15}$$

As in the first example, it follows that the canonical $\mathfrak{sl}_2$-triple associated to $\mathbb{O}_{sr}$ is

$$\begin{aligned}
e &= X_{\alpha_1} + X_{\alpha_1+\alpha_2} - X_{\alpha_2+\alpha_4} + 2X_{\alpha_3} - X_{\alpha_4}, \\
h &= 4H_{\alpha_1} + 6H_{\alpha_2} + 4H_{\alpha_3} + 4H_{\alpha_4}, \\
f &= X_{-\alpha_1} + 3X_{-\alpha_1-\alpha_2} - 3X_{-\alpha_2-\alpha_4} + 2X_{-\alpha_3} - X_{-\alpha_4}.
\end{aligned}$$

We can now define the basis vectors Z_i of $\mathfrak{g}(e)$ and X_j of an ad_h-invariant complementary subspace $\mathfrak{n}$ to $\mathfrak{g}(e)$ in the Chevalley basis:

$$\begin{aligned}
Z_1 &= X_{\alpha_1+\alpha_2} - X_{\alpha_2+\alpha_4} + 2X_{\alpha_3}, \\
Z_2 &= X_{\alpha_1+\alpha_2+\alpha_3+\alpha_4}, \\
Z_3 &= X_{\alpha_1+2\alpha_2+\alpha_3+\alpha_4}, \\
Z_4 &= X_{\alpha_1} - X_{\alpha_4}, \\
Z_5 &= X_{\alpha_2+\alpha_3} + X_{\alpha_2+\alpha_4} - X_{\alpha_3} - X_{\alpha_4}, \\
Z_6 &= X_{\alpha_1+\alpha_2+\alpha_3} + X_{\alpha_1+\alpha_2+\alpha_4} - X_{\alpha_2+\alpha_3+\alpha_4},
\end{aligned} \tag{16}$$

$$\begin{array}{lll}
X_1 = X_{\alpha_1+\alpha_2+\alpha_3}, & X_8 = H_{\alpha_3}, & X_{15} = X_{-\alpha_1-\alpha_2}, \\
X_2 = X_{\alpha_2+\alpha_3+\alpha_4}, & X_9 = H_{\alpha_4}, & X_{16} = X_{-\alpha_2-\alpha_3}, \\
X_3 = X_{\alpha_4}, & X_{10} = X_{\alpha_2}, & X_{17} = -X_{-\alpha_2-\alpha_4}, \\
X_4 = X_{\alpha_3}, & X_{11} = X_{-\alpha_2}, & X_{18} = -X_{-\alpha_1-\alpha_2-\alpha_3}, \\
X_5 = X_{\alpha_2+\alpha_4}, & X_{12} = X_{-\alpha_1}, & X_{19} = -X_{-\alpha_1-\alpha_2-\alpha_4}, \\
X_6 = H_{\alpha_1}, & X_{13} = X_{-\alpha_3}, & X_{20} = -X_{-\alpha_2-\alpha_3-\alpha_4}, \\
X_7 = H_{\alpha_2}, & X_{14}, = -X_{-\alpha_4}, & X_{21} = -X_{-\alpha_1-\alpha_2-\alpha_3-\alpha_4}, \\
 & & X_{22} = -X_{-\alpha_1-2\alpha_2-\alpha_3-\alpha_4}.
\end{array}$$

If we denote by $\bar{Z}_1, \dots, \bar{Z}_6$ the dual basis (with respect to $\langle X \mid Y \rangle = \frac{1}{2}\,\mathrm{Trace}(XY)$) of the basis $Z_1, \dots, Z_6$ of $\mathfrak{g}(e)$, then a typical element of the transverse slice $N =$

$e+\mathfrak{n}^\perp$ is $e+\sum_{i=1}^{6} q_i \bar{Z}_i$, that is,

$$Q=\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0\\ q_4 & 0 & 0 & 0 & 0 & 0 & 0 & -1\\ q_1 & 0 & 0 & 2 & 0 & 0 & 0 & -1\\ 0 & q_5 & 0 & 0 & 0 & 1 & 1 & 0\\ 0 & -q_3 & -q_2 & 0 & 0 & -q_4 & -q_1 & 0\\ q_3 & 0 & q_6 & 0 & -1 & 0 & 0 & -q_5\\ q_2 & -q_6 & 0 & 0 & -1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 \end{pmatrix}, \tag{17}$$

and we can compute the matrix A restricted to N by $A_{ij}=\langle Q \mid [Z_i, Z_j]\rangle$, and similarly for the matrices B and C. A direct substitution in (4) leads to the Poisson matrix for the ATP-structure:

$$\Lambda_N=\frac{1}{2}\begin{pmatrix} 0 & q_4q_6 & -q_4q_6 & 0 & -2q_6 & 2q_{16}\\ -q_4q_6 & 0 & 0 & q_4q_6 & -q_5q_6 & -2q_{36}\\ q_4q_6 & 0 & 0 & -q_4q_6 & q_5q_6 & 2q_{36}\\ 0 & -q_4q_6 & q_4q_6 & 0 & 2q_6 & -2q_{16}\\ 2q_6 & q_5q_6 & -q_5q_6 & -2q_6 & 0 & 2q_{56}\\ -2q_{16} & 2q_{36} & -2q_{36} & 2q_{16} & -2q_{56} & 0 \end{pmatrix}, \tag{18}$$

where

$$\begin{aligned} q_{16}&=2q_2-q_1q_4-q_4q_5+q_4^2,\\ q_{36}&=q_3q_4-q_2q_4-q_2q_5,\\ q_{56}&=2q_3-2q_2-q_5^2+q_4q_5-q_1q_5. \end{aligned} \tag{19}$$

It follows from (15) and (16) that the quasidegrees of the variables q_i are $\varpi(q_1)=\varpi(q_4)=\varpi(q_5)=4$, $\varpi(q_2)=\varpi(q_3)=8$, and $\varpi(q_6)=6$. That the ATP-structure is quasihomogeneous of quasidegree -2 can again be easily read off from (18).

5. The subregular case

In this section we will explicitly describe the ATP-structure of the subregular orbit $\mathcal{O}_{sr}\subset\mathfrak{g}$, where $\mathfrak{g}$ is a semisimple Lie algebra. Since in the subregular orbit the generic rank of the ATP-structure on the transverse slice N is two, and since we know $\dim(N)-2$ independent Casimirs, namely the basic Ad-invariant functions on $\mathfrak{g}$ restricted to N, we will easily derive that the ATP-structure is the determinantal structure (also called Nambu structure) determined by these Casimirs, up to multiplication by a *function*. What is much less trivial to show is that this function is only a *constant*. For this we will use Brieskorn's theory of simple singularities, which is recalled in Section 5 below. First we recall the basic facts on Ad-invariant functions on $\mathfrak{g}$ and link them to the ATP-structure.

Invariant functions and Casimirs. Let $\mathbb{O}_{sr} = \mathrm{G} \cdot e$, be a subregular orbit in the semisimple Lie algebra $\mathfrak{g}$. Let (h, e, f) be the corresponding canonical $\mathfrak{sl}_2$-triple, and consider the transverse slice $N := e + \mathfrak{n}^{\perp}$ to $\mathrm{G} \cdot e$, where $\mathfrak{n}$ is an ad_h-invariant complement to $\mathfrak{g}(e)$. We know from Section 3 that the ATP-structure on N, when equipped with the linear coordinates $q_1, \dots, q_k$, is a quasihomogeneous polynomial Poisson structure of generic rank 2. Let $S(\mathfrak{g}^*)^{\mathrm{G}}$ be the algebra of Ad-invariant polynomial functions on $\mathfrak{g}$. By a classical theorem due to Chevalley, $S(\mathfrak{g}^*)^{\mathrm{G}}$ is a polynomial algebra generated by ℓ homogeneous polynomials $(G_1, \dots, G_\ell)$ whose degree $d_i := \deg(G_i) = m_i + 1$, where $m_1, \dots, m_\ell$ are the exponents of $\mathfrak{g}$. These functions are Casimirs of the Lie–Poisson structure on $\mathfrak{g}$, since Ad-invariance of G_i implies that $[x, dG_i(x)] = 0$, and hence the Lie–Poisson bracket (2) is

$$\{F, G_i\}(x) = \langle x \mid [dF(x), dG_i(x)]\rangle = -\langle [x, dG_i(x)] \mid dF(x)\rangle = 0$$

for any function F on $\mathfrak{g}$. If we denote by χ_i the restriction of G_i to the transverse slice N then, it follows that these functions are Casimirs of the ATP-structure. The polynomials χ_i are not homogeneous, but they are quasihomogeneous.

Lemma 5.1. *Each χ_i is a quasihomogeneous polynomial of quasidegree $2d_i$ relative to the quasidegrees $(2+n_1, \dots, 2+n_k)$.*

Proof. Since χ_i is of degree d_i and χ_i is Ad-invariant, we get

$$\rho_t^*(\chi_i) = \chi_i \circ \rho_{t^{-1}} = \chi_i \circ (t^{-2}\,\mathrm{Ad}_{\lambda^{-1}(t)}) = t^{-2d_i}\chi_i \circ \mathrm{Ad}_{\lambda^{-1}(t)} = t^{-2d_i}\chi_i,$$

so that χ_i has quasidegree $2d_i$. $\square$

Simple singularities. Let $\mathfrak{h}$ be a Cartan subalgebra of $\mathfrak{g}$. The Weyl group $\mathcal{W}$ acts on $\mathfrak{h}$, and the algebra $S(\mathfrak{g}^*)^{\mathrm{G}}$ of Ad-invariant polynomial functions on $\mathfrak{g}$ is isomorphic to $S(\mathfrak{h}^*)^{\mathcal{W}}$, the algebra of $\mathcal{W}$-invariant polynomial functions on $\mathfrak{h}^*$. The inclusion homomorphism $S(\mathfrak{g}^*)^{\mathrm{G}} \hookrightarrow S(\mathfrak{g}^*)$, is dual to a morphism $\mathfrak{g} \to \mathfrak{h}/\mathcal{W}$, called the *adjoint quotient*. Concretely, the adjoint quotient is given by

$$\begin{aligned} G : \mathfrak{g} &\to \mathbb{C}^\ell \\ x &\mapsto (G_1(x),\ G_2(x), \dots, G_\ell(x)). \end{aligned} \tag{20}$$

The zero-fiber $G^{-1}(0)$ of G is exactly the nilpotent variety $\mathcal{N}$ of $\mathfrak{g}$. As we are interested in $N \cap \mathcal{N} = N \cap G^{-1}(0) = \chi^{-1}(0)$ — which is an affine surface with an isolated, simple singularity — let us recall the notion of simple singularity (see [Slodowy 1980a] for details). Up to conjugacy, there are five types of finite subgroups of $\mathrm{SL}_2 = \mathrm{SL}_2(\mathbb{C})$, which are denoted by $\mathcal{C}_p$, $\mathcal{D}_p$, $\mathcal{T}$, $\mathbb{O}$, and $\mathcal{I}$. Given such a subgroup $\boldsymbol{F}$, one looks at the corresponding ring of invariant polynomials $\mathbb{C}[u, v]^{\boldsymbol{F}}$.

In each of the five cases, $\mathbb{C}[u, v]^{\boldsymbol{F}}$ is generated by three fundamental polynomials X, Y, Z, subject to only one relation $R(X, Y, Z) = 0$; hence the quotient space $\mathbb{C}^2/\boldsymbol{F}$ can be identified, as an affine surface, with the singular surface in $\mathbb{C}^3$ defined by $R = 0$. The origin is its only singular point; it is called a *(homogeneous) simple singularity*. The exceptional divisor of the minimal resolution of $\mathbb{C}^2/\boldsymbol{F}$ is a finite set of projective lines. If two of these lines meet, then they meet in a single point, and transversally. Moreover, the intersection pattern of these lines forms a graph that coincides with one of the simply laced Dynkin diagrams of type A_ℓ, D_ℓ, E_6, E_7, or E_8. This type is called the type of the singularity. Moreover, every such Dynkin diagram (that is, of type ADE) appears in this way; see Table 1.

For the other simple Lie algebras (of type B_ℓ, C_ℓ, F_4 or G_2), there exists a similar correspondence. By definition, an *(inhomogeneous) simple singularity* of type Δ is a couple (V, Γ) consisting of a homogeneous simple singularity $V = \mathbb{C}^2/\boldsymbol{F}$ and a group $\Gamma = \boldsymbol{F}'/\boldsymbol{F}$ of automorphisms of V, according to Table 2.

The connection between the diagram of (V, Γ) and that of V can be described as follows. The action of Γ on V lifts to an action on a minimal resolution of V that permutes the components of the exceptional set. Then, we obtain the diagram of (V, Γ) as a Γ-quotient of that of V. It leads to Table 3, which is the nonsimply-laced analog of Table 1.

We can now state an extension of a theorem of Brieskorn.

Proposition 5.2 [Slodowy 1980a, Theorems 1 and 2]. *Let $\mathfrak{g}$ be a simple complex Lie algebra with Dynkin diagram of type Δ. Let $\mathbb{O}_{sr} = \mathrm{G}\cdot e$ be the subregular orbit, and let $N = e + \mathfrak{n}^\perp$ be a transverse slice to $\mathrm{G}\cdot e$. The surface $N \cap \mathcal{N} = \chi^{-1}(0)$ has a (homogeneous or inhomogeneous) simple singularity of type Δ.*

To finish this section we illustrate the results above for the examples of Section 4. In both cases we give the invariants restricted to the slice N and their zero locus,

Group $\boldsymbol{F}$	Singularity $R(X, Y, Z) = 0$	Type Δ
$\mathscr{C}_{\ell+1}$	$X^{\ell+1} + YZ = 0$	A_ℓ
$\mathscr{D}_{\ell-2}$	$X^{\ell-1} + XY^2 + Z^2 = 0$	D_ℓ
$\mathscr{T}$	$X^4 + Y^3 + Z^2 = 0$	E_6
$\mathbb{O}$	$X^3Y + Y^3 + Z^2 = 0$	E_7
$\mathscr{I}$	$X^5 + Y^3 + Z^2 = 0$	E_8

Table 1. The basic correspondence between finite subgroups $\boldsymbol{F}$ of SL_2, homogeneous simple singularities defined by an equation $R(X, Y, Z) = 0$, and simply laced simple Lie algebras of type Δ.

the surface $\chi^{-1}(0)$.

First, for the subregular orbit of $\mathfrak{g}_2$, the invariant functions restricted to the slice N are

$$\begin{aligned}\chi_1 &= q_1,\\ \chi_2 &= 12q_1q_2q_3 - 4q_2^3 - 4q_3^3 + 9q_4^2,\end{aligned} \tag{21}$$

which leads to an affine surface $\chi^{-1}(0)$ in $\mathbb{C}^4$ that is isomorphic to the surface in $\mathbb{C}^3$ defined by

$$4q_2^3 + 4q_3^3 - 9q_4^2 = 0.$$

Up to a rescaling, this is the polynomial R that was given in Table 3.

Second, for the subregular orbit of $\mathfrak{so}_8$, the invariant functions restricted to the slice N are found as the (nonconstant) coefficients of the characteristic polynomial

Type Δ	V	$\boldsymbol{F}$	$\boldsymbol{F}'$	$\Gamma = \boldsymbol{F}'/\boldsymbol{F}$
B_ℓ	$A_{2\ell-1}$	$\mathscr{C}_{2\ell}$	$\mathscr{D}_\ell$	$\mathbb{Z}/2\mathbb{Z}$
C_ℓ	$D_{\ell+1}$	$\mathscr{D}_{\ell-1}$	$\mathscr{D}_{2\ell-2}$	$\mathbb{Z}/2\mathbb{Z}$
F_4	E_6	$\mathscr{T}$	$\mathbb{O}$	$\mathbb{Z}/2\mathbb{Z}$
G_2	D_4	$\mathscr{D}_2$	$\mathbb{O}$	$\mathbb{Z}/3\mathbb{Z}$

Table 2. List of all possible inhomogeneous singularities of type $\Delta = (V, \Gamma)$, where V is one of the homogeneous simple singularities and $\Gamma = \boldsymbol{F}'/\boldsymbol{F}$ is a group of automorphisms of V. The labels B_ℓ, C_ℓ, F_4 and G_2 for these types will become clear in Proposition 5.2.

Type Δ	Singularity $R(X, Y, Z) = 0$	Γ-action
B_ℓ	$X^{2\ell} + YZ = 0$	$(X, Y, Z) \longrightarrow (-X, Z, Y)$
C_ℓ	$X^\ell + XY^2 + Z^2 = 0$	$(X, Y, Z) \longrightarrow (X, -Y, -Z)$
F_4	$X^4 + Y^3 + Z^2 = 0$	$(X, Y, Z) \longrightarrow (-X, Y, -Z)$
G_2	$X^3 + Y^3 + Z^2 = 0$	$(X, Y, Z) \longrightarrow (\alpha X, \alpha^2 Y, Z)$

Table 3. For each of the inhomogeneous simple singularities of type Δ (see Table 2), the corresponding homogeneous simple singularity $V = \mathbb{C}^2/\boldsymbol{F}$ is given by its equation $R(X, Y, Z) = 0$ together with the action of $\Gamma = \boldsymbol{F}'/\boldsymbol{F}$ on V. In the last line, α is a nontrivial cubic root of unity.

of the matrix Q (see (17)):

$$
\begin{aligned}
\chi_1 &= -2q_1 - 2q_4,\\
\chi_2 &= -12q_2 - 4q_3 - 4q_4q_5 + (q_1+q_4)^2,\\
\chi_3 &= -q_2 + q_3 - q_4q_5,\\
\chi_4 &= -4q_1q_2 - 16q_2q_5 - 12q_3q_4 + 12q_2q_4 + 4q_1q_3 + 4q_4^2q_5 + 4q_1q_4q_5 - 4q_6^2.
\end{aligned}
\tag{22}
$$

By linearly eliminating the variables q_1, q_2 and q_3 from the equations $\chi_i = 0$ for $i = 1, 2, 3$, we find that $\chi^{-1}(0)$ is isomorphic to the affine surface in $\mathbb{C}^3$ defined by

$$4q_4^2q_5 - 2q_4q_5^2 + q_6^2 = 0.$$

Its defining polynomial corresponds to the polynomial R in Table 1, after putting $X = i\gamma q_4$, $Y = \gamma(q_5 - q_4)$, and $Z = q_6$, where γ is any cubic root of $2i$.

The determinantal Poisson structure. We prove here the announced result that the ATP-structure in the subregular case is a determinantal Poisson structure determined by the Casimirs. Let us first point out how such a structure is defined. Let $C_1, \dots, C_{d-2}$ be $d-2$ (algebraically) independent polynomials in $d > 2$ variables $x_1, \dots, x_d$. For a polynomial F in the variables $x_1, \dots, x_d$, let us denote by ∇F its differential $\mathrm{d}F$, expressed in the natural basis $\mathrm{d}x_i$, that is, ∇F is a column vector with elements $(\nabla F)_i = \partial F/\partial x_i$. Then a polynomial Poisson structure is defined on $\mathbb{C}^d$ by

$$\{F, G\}_{\det} := \det(\nabla F,\ \nabla G,\ \nabla C_1,\ \dots,\ \nabla C_{d-2}), \tag{23}$$

where F and G are arbitrary polynomials. It is clear that each of the C_i is a Casimir of $\{\cdot,\cdot\}_{\det}$, so that in particular the generic rank of $\{\cdot,\cdot\}_{\det}$ is two. Notice also that if the Casimirs C_i are quasihomogeneous with respect to the weights $\varpi_i := \varpi(x_i)$, then for any quasihomogeneous elements F and G we have

$$\varpi(\{F, G\}_{\det}) = \varpi(F) + \varpi(G) + \sum_{i=1}^{d-2} \varpi(C_i) - \sum_{i=1}^{d} \varpi_i.$$

This follows easily from the definition of a determinant and that if F is any quasihomogeneous polynomial, then $\partial F/\partial x_i$ is quasihomogeneous and $\varpi(\partial F/\partial x_i) = \varpi(F) - \varpi_i$. It follows that $\{\cdot,\cdot\}_{\det}$ is quasihomogeneous of quasidegree κ, where

$$\kappa = \sum_{i=1}^{d-2} \varpi(C_i) - \sum_{i=1}^{d} \varpi_i. \tag{24}$$

Applied to our case, it means that we have two polynomial Poisson structures on the transverse slice N that have $\chi_1, \dots, \chi_\ell$ as Casimirs on $N \cong \mathbb{C}^{\ell+2}$, namely, the ATP-structure and the determinantal structure constructed by using these Casimirs.

Remark 5.3. The determinantal Poisson structure first appears (without proof that it is a Poisson structure) in [Damianou 1989], who attributes the formula to H. Flaschka and T. Ratiu. The first explicit proof appears in [Grabowski et al. 1993]. A more conceptual proof appears in [Takhtajan 1994, Remark 1 and Theorem 4].

In our two examples (see Section 4), these structures are easily compared by explicit computation. For the subregular orbit of $\mathfrak{g}_2$, we have, according to (23),

$$(\Lambda_{\det})_{ij} = \det\left(\nabla q_i \ \nabla q_j \ \nabla\chi_1 \ \nabla\chi_2\right),$$

where χ_1 and χ_2 are the Casimirs (21). This leads to

$$\Lambda_{\det} = -6\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -3q_4 & 2q_1q_2-2q_3^2 \\ 0 & 3q_4 & 0 & 2q_2^2-2q_1q_3 \\ 0 & -2q_1q_2+2q_3^2 & -2q_2^2+2q_1q_3 & 0 \end{pmatrix}.$$

In view of (14), it follows that $\Lambda_{\det} = -6\Lambda_N$, so that both Poisson structures coincide. For $\mathfrak{so}_8$, one finds similarly, using the Casimirs $\chi_1, \dots, \chi_4$ in (22),

$$\Lambda_{\det} = -128\begin{pmatrix} 0 & q_4q_6 & -q_4q_6 & 0 & -2q_6 & 2q_{16} \\ -q_4q_6 & 0 & 0 & q_4q_6 & -q_5q_6 & -2q_{36} \\ q_4q_6 & 0 & 0 & -q_4q_6 & q_5q_6 & 2q_{36} \\ 0 & -q_4q_6 & q_4q_6 & 0 & 2q_6 & -2q_{16} \\ 2q_6 & q_5q_6 & -q_5q_6 & -2q_6 & 0 & 2q_{56} \\ -2q_{16} & 2q_{36} & -2q_{36} & 2q_{16} & -2q_{56} & 0 \end{pmatrix},$$

where q_{16}, q_{36} and q_{56} are given by (19). In view of (18), both Poisson structures again coincide, $\Lambda_{\det} = -256\Lambda_N$.

To show that, in the subregular case, the ATP-structure and the determinantal structure always coincide, that is, they differ only by a constant factor, we first show that both structures coincide up to a rational function.

Proposition 5.4. *Suppose $\{\cdot,\cdot\}$ and $\{\cdot,\cdot\}'$ are two nontrivial polynomial Poisson structures on $\mathbb{C}^d$ that have $d-2$ common independent polynomial Casimirs $C_1, \dots, C_{d-2}$. Then there exists a rational function $R \in \mathbb{C}(x_1, \dots, x_d)$ such that $\{\cdot,\cdot\} = R\{\cdot,\cdot\}'$.*

Proof. Let M and M' denote the Poisson matrices of $\{\cdot,\cdot\}$ and $\{\cdot,\cdot\}'$ in the coordinates $x_1, \dots, x_d$. If we denote $\mathcal{R} := \mathbb{C}(x_1, \dots, x_d)$, then M and M' both act naturally as skew-symmetric endomorphisms on the $\mathcal{R}$-vector space $\mathcal{R}^d$. The subspace H of $\mathcal{R}^d$ spanned by $\nabla C_1, \dots, \nabla C_{d-2}$ is the kernel of both maps; hence we have two induced skew-symmetric endomorphisms φ and φ' of the quotient space $\mathcal{R}^d/H$. Since the latter is two-dimensional, φ' and φ are proportional, that is, $\varphi' = R\varphi$ with $R \in \mathcal{R}$. Since M and M' have the same kernel, $M' = RM$. $\square$

Applied to our two Poisson structures $\{\cdot\,,\cdot\}_N$ and $\{\cdot\,,\cdot\}_{\mathrm{det}}$, the proposition yields that $\{\cdot\,,\cdot\}_N = R\,\{\cdot\,,\cdot\}_{\mathrm{det}}$, where $R = P/Q \in \mathcal{R}$. We show next that R is actually a (nonzero) constant and thereby characterize completely the ATP-structure in the subregular case.

Theorem 5.5. *Let $\mathbb{O}_{sr}$ be the subregular nilpotent adjoint orbit of a complex semisimple Lie algebra $\mathfrak{g}$, and let (h, e, f) be the canonical triple associated to $\mathbb{O}_{sr}$. Let $N = e + \mathfrak{n}^\perp$ be a slice transverse to $\mathbb{O}_{sr}$, where $\mathfrak{n}$ is an ad_h-invariant complementary subspace to $\mathfrak{g}(e)$. Let $\{\cdot\,,\cdot\}_N$ and $\{\cdot\,,\cdot\}_{det}$ denote respectively the ATP-structure and the determinantal structure on N. Then $\{\cdot\,,\cdot\}_N = c\,\{\cdot\,,\cdot\}_{det}$ for some $c \in \mathbb{C}^*$.*

Proof. By the above, $\{\cdot\,,\cdot\}_N = R\,\{\cdot\,,\cdot\}_{\mathrm{det}}$, where $R \in \mathcal{R}$. If R has a nontrivial denominator Q, then all elements of the Poisson matrix of $\{\cdot\,,\cdot\}_{\mathrm{det}}$ must be divisible by Q, since both Poisson structures are polynomial. Then along the hypersurface $Q = 0$, the rank of $(\nabla\chi_1, \dots, \nabla\chi_\ell)$ is smaller than ℓ; hence $\chi^{-1}(0)$ is singular along the curve $\chi^{-1}(0) \cap (Q = 0)$. However, by Proposition 5.2, we know that $\chi^{-1}(0)$ has an isolated singularity, which leads to a contradiction. This shows that Q is a constant and hence that R is a polynomial.

To show that the polynomial R is constant, it suffices to show that the quasidegrees of $\{\cdot\,,\cdot\}_N$ and $\{\cdot\,,\cdot\}_{\mathrm{det}}$ are the same, which amounts (in view of Proposition 3.2) to showing that the quasidegree of $\{\cdot\,,\cdot\}_{\mathrm{det}}$ is -2. This follows from the following formula due to Kostant [1963, Thm 7], which expresses the dimension of the regular orbit in terms of the exponents m_i of $\mathfrak{g}$:

$$2\sum_{i=1}^{\ell} m_i = \dim \mathbb{O}_{reg} = \dim \mathfrak{g} - \ell.$$

Indeed, if we apply this formula, Lemma 5.1, and (8) to the formula (24) for the quasidegree of $\{\cdot\,,\cdot\}_{\mathrm{det}}$, then we find

$$\begin{aligned}\kappa &= \sum_{i=1}^{\ell} \varpi(\chi_i) - \sum_{i=1}^{\ell+2} \varpi(q_i) = 2\sum_{i=1}^{\ell} d_i - \sum_{i=1}^{\ell+2}(n_i + 2)\\ &= 2\sum_{i=1}^{\ell} m_i - \sum_{i=1}^{\ell+2} n_i - 4\\ &= \dim \mathfrak{g} - \ell - (\dim \mathfrak{g} - \ell - 2) - 4 = -2.\end{aligned}$$

$\square$

Reduction to a 3×3 Poisson matrix. Let $\mathbb{O}_{sr}$ be the subregular nilpotent adjoint orbit of a complex semisimple Lie algebra $\mathfrak{g}$ of rank ℓ. Let (h, e, f) be its associated canonical $\mathfrak{sl}_2$-triple, and let $N := e + \mathfrak{n}^{\perp}$ be a transverse slice to $\mathbb{O}_{sr}$, where $\mathfrak{n}$ is an ad_h-invariant complementary subspace to $\mathfrak{g}(e)$. Let $\{\cdot\,,\cdot\}_N$ be the ATP-structure defined on N. Recall that N is equipped with linear coordinates $q_1, \dots, q_{\ell+2}$ defined in Section 2, and that $\{\cdot\,,\cdot\}_N$ has independent Casimirs $\chi_1, \dots, \chi_\ell$, which are the restrictions to N of the basic homogeneous invariant polynomial functions on $\mathfrak{g}$.

Our goal now is to show that, in well-chosen coordinates, the ATP-structure $\{\cdot\,,\cdot\}_N$ on N is essentially given by a 3×3 skew-symmetric matrix which is closely related to the polynomial that defines the singularity. More precisely:

Theorem 5.6. *After possibly relabeling the coordinates q_i and the Casimirs χ_i, the $\ell + 2$ functions*

$$\chi_i,\ 1 \le i \le \ell - 1, \quad \textit{and} \quad q_\ell,\ q_{\ell+1},\ q_{\ell+2}$$

form a system of (global) coordinates on the affine space N. The Poisson matrix of the ATP-structure on N in these coordinates is

$$\widetilde{\Lambda}_N = \begin{pmatrix} 0 & 0 \\ 0 & \Omega \end{pmatrix}, \quad \textit{where} \quad \Omega = c' \begin{pmatrix} 0 & \dfrac{\partial \chi_\ell}{\partial q_{\ell+2}} & -\dfrac{\partial \chi_\ell}{\partial q_{\ell+1}} \\ -\dfrac{\partial \chi_\ell}{\partial q_{\ell+2}} & 0 & \dfrac{\partial \chi_\ell}{\partial q_\ell} \\ \dfrac{\partial \chi_\ell}{\partial q_{\ell+1}} & -\dfrac{\partial \chi_\ell}{\partial q_\ell} & 0 \end{pmatrix}, \tag{25}$$

for some nonzero constant c'. It has the polynomial χ_ℓ as Casimir, which reduces to the polynomial that defines the singularity if we set $\chi_j = 0$ for $j = 1, 2, \dots, \ell - 1$.

Proof. The non-Poisson part of this theorem is due to Brieskorn and Slodowy. Before proving the Poisson part of the theorem, namely, that the Poisson matrix takes the form (25), we explain for the reader's convenience the basics of singularity theory used in their proof, but see [Slodowy 1980a] for details. Let (X_0, x) be the germ of an analytic variety X_0 at the point x. A deformation of (X_0, x) is a pair $(\Phi, \imath)$ where $\Phi : X \to U$ is a flat morphism of varieties with $\Phi(x) = u$ and where the map $\imath : X_0 \to \Phi^{-1}(u)$ is an isomorphism. Such a deformation is called *semiuniversal* if any other deformation of (X_0, x) is isomorphic to a deformation induced from $(\Phi, \imath)$ by a local change of variables in a neighborhood of x. The semiuniversal deformation of (X_0, x) is unique up to isomorphism. It can be explicitly described in the following case. Let $(X_0, 0)$ be a germ of a hypersurface of $\mathbb{C}^d$ that is singular at 0, and say X_0 is locally given by $f(z) = 0$. Then the semiuniversal deformation of $(X_0, 0)$ is the (germ at the origin of the) map

$$\begin{array}{rcl} \Phi \,:\, \mathbb{C}^k \times \mathbb{C}^d & \to & \mathbb{C}^k \times \mathbb{C} \\ (u, z) & \mapsto & (u, F(u, z)), \end{array}$$

where

$$F(u, z) = f(z) + \sum_{i=1}^{k} g_i(z) u_i$$

and where the polynomials $1, g_1, g_2, \dots, g_k$ represent a vector space basis of the Milnor (or Tjurina) algebra

$$\mathcal{M}(f) := \frac{\mathbb{C}[z_1, \dots, z_d]}{\left(f, \dfrac{\partial f}{\partial z_1}, \dots, \dfrac{\partial f}{\partial z_d}\right)} = \frac{\mathbb{C}[z_1, \dots, z_d]}{\left(\dfrac{\partial f}{\partial z_1}, \dots, \dfrac{\partial f}{\partial z_d}\right)}. \tag{26}$$

The last equality is valid whenever f is quasihomogeneous, which is true in this case. The dimension $\dim \mathcal{M}(f) = k + 1$ is called the *Milnor number* of f.

We can now formulate Brieskorn's result. It says that the map $\chi : N \to \mathbb{C}^\ell$, which is the restriction of the adjoint quotient (20) to the slice N, is a semiuniversal deformation of the singular surface $N \cap \mathcal{N}$. More precisely, when the Lie algebra is of the type ADE, then the map

$$\begin{array}{rccl} \Phi \,: & \mathbb{C}^{\ell-1} \times \mathbb{C}^3 & \to & \mathbb{C}^{\ell-1} \times \mathbb{C} \\ & ((\chi_1, \dots, \chi_{\ell-1}), (q_\ell, q_{\ell+1}, q_{\ell+2})) & \mapsto & ((\chi_1, \dots, \chi_{\ell-1}), \chi_\ell) \end{array}$$

is the semiuniversal deformation of the singular surface $N \cap \mathcal{N}$; for the other types one has to consider Γ-invariant semiuniversal deformations, as was shown by Slodowy [1980a], see Table 2. It is implicit in Brieskorn's statement that $(\chi_1, \dots, \chi_{\ell-1}, q_\ell, q_{\ell+1}, q_{\ell+2})$ form a system of coordinates on N, which comes from the fact that one can solve the $\ell - 1$ equations $\chi_i = \chi_i(q)$ *linearly* for $\ell - 1$ of the variables q_i. That is, the Casimirs have the form

$$\begin{pmatrix} \chi_1 \\ \vdots \\ \chi_{\ell-1} \end{pmatrix} = A \begin{pmatrix} q_1 \\ \vdots \\ q_{\ell-1} \end{pmatrix} + \begin{pmatrix} F_1(q_\ell, q_{\ell+1}, q_{\ell+2}) \\ \vdots \\ F_{\ell-1}(q_\ell, q_{\ell+1}, q_{\ell+2}) \end{pmatrix}, \tag{27}$$

where A is a constant matrix with $\det A \neq 0$; this will be illustrated in the examples below.

We now get to the Poisson part of the proof. Since the coordinate functions $\chi_1, \dots, \chi_{\ell-1}$ are Casimirs, the Poisson matrix $\widetilde{\Lambda}_N$ has with respect to these coordinates the block form

$$\widetilde{\Lambda}_N = \begin{pmatrix} 0 & 0 \\ 0 & \Omega \end{pmatrix},$$

where Ω is a 3×3 skew-symmetric matrix. We know from Theorem 5.5 that the ATP-structure is a constant multiple of the determinantal structure. Since $\det A$

lies in $\mathbb{C}^*$, it follows from (27) that, for $\ell \le i,\ j \le \ell + 2$,

$$\widetilde{\Lambda}_{ij} := c \det(\nabla q_i \ \nabla q_j \ \nabla \chi_1 \ \ldots \ \nabla \chi_\ell) = c' \det(\nabla' q_i \ \nabla' q_j \ \nabla' \chi_\ell),$$

where c and c' are nonzero constants and ∇' denotes the restriction of ∇ to $\mathbb{C}^3$, namely,

$$\nabla' F = \left(\frac{\partial F}{\partial q_\ell} \ \frac{\partial F}{\partial q_{\ell+1}} \ \frac{\partial F}{\partial q_{\ell+2}} \right)^\top .$$

The explicit formula (25) for Ω follows at once. □

6. Examples

The subregular orbit of $\mathfrak{g}_2$. For this we have, according to (21), that $\chi_1 = q_1$. Then χ_2 expressed in terms of q_2, q_3, q_4, and χ_1 is

$$\chi_2 = 9q_4^2 - 4q_2^3 - 4q_3^3 + 12\chi_1 q_2 q_3.$$

The Poisson matrix (14) of the ATP-structure is already in the form (25), with $c' = -1/6$ (and $\chi_1 = q_1$). Since the Milnor algebra (26) is given in this case by $\mathcal{M}(9q_4^2 - 4q_2^3 - 4q_3^3) = \mathbb{C}[q_2, q_3, q_4] / \left(q_2^2, q_3^2, q_4\right)$, one easily sees that 1 and the coefficient $q_2 q_3$ of u_1 indeed form a vector space basis for the Γ-invariant elements of the Milnor algebra (see Table 3); compare [Slodowy 1980a, page 136].

The subregular orbit of $\mathfrak{so}_8$. Recall from (22) that its ATP structure has Casimirs $\chi_1, \ldots, \chi_4$. As stated in the proof of Theorem 5.6, we can solve three of them linearly for q_1, q_2, q_3 in terms of χ_1, χ_2, χ_3 and the last three variables q_4, q_5, and q_6. We obtain

$$\begin{aligned}
q_1 &= -q_4 - \tfrac{1}{2}\chi_1, \\
q_2 &= \tfrac{1}{64}\left(\chi_1^2 - 16\chi_3 - 4\chi_2 - 32 q_4 q_5\right), \\
q_3 &= \tfrac{1}{64}\left(\chi_1^2 + 48\chi_3 - 4\chi_2 + 32 q_4 q_5\right).
\end{aligned}$$

Substituted in χ_4, this yields

$$\chi_4 = 8q_4 q_5^2 - 16 q_4^2 q_5 - 4q_6^2 - 4\chi_1 q_4 q_5 + (\chi_2 - \tfrac{1}{4}\chi_1^2 + 4\chi_3) q_5 - 16\chi_3 q_4 - 2\chi_1\chi_3,$$

so that

$$\hat{\chi}_4 = 8q_4 q_5^2 - 16 q_4^2 q_5 - 4q_6^2 - 4\chi_1 q_4 q_5 + \hat{\chi}_2 q_5 - 16\chi_3 q_4,$$

where $\hat{\chi}_2 := \chi_2 - \frac{1}{4}\chi_1^2 + 4\chi_3$ and $\hat{\chi}_4 := \chi_4 + 2\chi_1\chi_3$ can be used instead of χ_2 and χ_4 as basic Ad-invariant polynomials restricted to N. Using (18), expressed in the coordinates $\chi_1, \hat{\chi}_2, \chi_3, q_4, q_5$ and q_6, we find that the matrix Ω is indeed of the form (25) with $c' = -\frac{1}{8}$, since

$$\{q_4, q_5\} = q_6 = -\frac{1}{8}\frac{\partial \hat{\chi}_4}{\partial q_6}, \quad \{q_4, q_6\} = 2q_4q_5 - 2q_4^2 - \frac{1}{2}\chi_1 q_4 + \frac{1}{8}\hat{\chi}_2 = \frac{1}{8}\frac{\partial \hat{\chi}_4}{\partial q_5},$$
$$\{q_5, q_6\} = -q_5^2 + 4q_4q_5 + \frac{1}{2}\chi_1 q_5 + 2\chi_3 = -\frac{1}{8}\frac{\partial \hat{\chi}_4}{\partial q_4}.$$

It follows easily that the Milnor algebra is given by

$$\mathcal{M}(8q_4q_5^2 - 16q_4^2q_5 - 4q_6^2) = \mathbb{C}[q_4, q_5, q_6]/(q_6, q_4(q_5 - q_4), q_5(q_5 - 4q_4)),$$

so that 1 and the coefficients q_4, q_5 and q_4q_5 of $\hat{\chi}_4$ indeed form a vector space basis for it.

The subregular orbit $\mathcal{O}_{sr}$ in $\mathfrak{sl}_4$. This example is from [Damianou 1996]. It was also examined by Sabourin [2005], who showed that the slice, originally due to Arnold [1971], belongs to the set $\mathcal{N}_h$. It is the orbit of the nilpotent element

$$e = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The transverse slice in Arnold's coordinates consists of matrices of the form

$$Q = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ q_1 & q_2 & q_3 & q_4 \\ q_5 & 0 & 0 & -q_3 \end{pmatrix}.$$

The basic Casimirs of the ATP-structure, as computed from the characteristic polynomial of Q, are

$$\chi_1 = q_2 + q_3^2, \quad \chi_2 = q_1 + q_2q_3, \quad \chi_3 = q_1q_3 + q_4q_5.$$

If we solve the first two equations for the variables q_1, q_2 in terms of χ_1, χ_2 and q_3, q_4, q_5, and substitute the result in χ_3, then we find that

$$\chi_3 = q_3^4 + q_4q_5 - \chi_1 q_3^2 + \chi_2 q_3.$$

Using the explicit formulas for the ATP-structure given in [Damianou 1996], expressed in the coordinates χ_1, χ_2, q_3, q_4 and q_5, we find that the matrix Ω is indeed of the form (25) with $c' = 1$, since

$$\{q_3, q_4\} = q_4 = \frac{\partial \chi_3}{\partial q_5}, \quad \{q_3, q_5\} = -q_5 = -\frac{\partial \chi_3}{\partial q_4},$$

$$\{q_4, q_5\} = 4q_3^3 - 2\chi_1 q_3 + \chi_2 = \frac{\partial \chi_3}{\partial q_3}.$$

It can be read from these formulas that the Milnor algebra is given by

$$\mathcal{M}(q_3^4 + q_4 q_5) = \mathbb{C}[q_3, q_4, q_5]/(q_4, q_5, q_3^3),$$

so that the coefficients 1, q_3 and q_3^2 of χ_3 indeed span its vector space.

References

[Arnold 1971] V. I. Arnold, "Matrices depending on parameters", *Uspehi Mat. Nauk* **26**:2(158) (1971), 101–114. In Russian; translated in *Russ. Math. Surv.* **26**:2 (1971), 29–43.

[Brieskorn 1971] E. Brieskorn, "Singular elements of semi-simple algebraic groups", pp. 279–284 in *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 2*, vol. 2, Gauthier-Villars, Paris, 1971. MR 55 #10720 Zbl 0223.22012

[Collingwood and McGovern 1993] D. H. Collingwood and W. M. McGovern, *Nilpotent orbits in semisimple Lie algebras*, Van Nostrand Reinhold Mathematics Series, Van Nostrand Reinhold Co., New York, 1993. MR 94j:17001 Zbl 0972.17008

[Cushman and Roberts 2002] R. Cushman and M. Roberts, "Poisson structures transverse to coadjoint orbits", *Bull. Sci. Math.* **126**:7 (2002), 525–534. MR 2003j:53126 Zbl 1074.53070

[Damianou 1989] P. A. Damianou, Ph.D. thesis, 1989.

[Damianou 1996] P. A. Damianou, "Transverse Poisson structures of coadjoint orbits", *Bull. Sci. Math.* **120**:2 (1996), 195–214. MR 97h:22010 Zbl 0853.58048

[Dixmier 1996] J. Dixmier, *Algèbres enveloppantes*, Les Grands Classiques Gauthier-Villars, Éditions Jacques Gabay, Paris, 1996. Reprint of the 1974 original. MR 98a:17012 Zbl 0867.17001

[Gan and Ginzburg 2002] W. L. Gan and V. Ginzburg, "Quantization of Slodowy slices", *Int. Math. Res. Not.* 5 (2002), 243–255. MR 2002m:53129 Zbl 0989.17014

[Grabowski et al. 1993] J. Grabowski, G. Marmo, and A. M. Perelomov, "Poisson structures: towards a classification", *Mod. Phys. Lett. A* **8**:18 (1993), 1719–1733. MR 94j:58063 Zbl 1020.37529

[Kostant 1963] B. Kostant, "Lie group representations on polynomial rings", *Amer. J. Math.* **85** (1963), 327–404. MR 28 #1252 Zbl 0124.26802

[Premet 2002] A. Premet, "Special transverse slices and their enveloping algebras", *Adv. Math.* **170**:1 (2002), 1–55. With an appendix by Serge Skryabin. MR 2003k:17014 Zbl 1005.17007

[Sabourin 2005] H. Sabourin, "Sur la structure transverse à une orbite nilpotente adjointe", *Canad. J. Math.* **57**:4 (2005), 750–770. MR 2006c:53089 Zbl 1077.22020

[Saint-Germain 1999] M. Saint-Germain, "Poisson algebras and transverse structures", *J. Geom. Phys.* **31**:2-3 (1999), 153–194. MR 2000m:53117 Zbl 0958.17006

[Slodowy 1980a] P. Slodowy, *Four lectures on simple groups and singularities*, vol. 11, Communications of the Mathematical Institute, Rijksuniversiteit Utrecht, Rijksuniversiteit Utrecht Mathematical Institute, Utrecht, 1980. MR 82b:14002 Zbl 0425.22020

[Slodowy 1980b] P. Slodowy, *Simple singularities and simple algebraic groups*, Lecture Notes in Mathematics **815**, Springer, Berlin, 1980. MR 82g:14037 Zbl 0441.14002

[Takhtajan 1994] L. Takhtajan, "On foundation of the generalized Nambu mechanics", *Comm. Math. Phys.* **160**:2 (1994), 295–315. MR 95b:58067 Zbl 0808.70015

[Tauvel 1998] P. Tauvel, *Introduction à la théorie des algèbres de Lie*, Diderot, Paris, 1998.

[Tauvel and Yu 2005] P. Tauvel and R. W. T. Yu, *Lie algebras and algebraic groups*, Springer Monographs in Mathematics, Springer, Berlin, 2005. MR 2006c:17001 Zbl 1068.17001

[Weinstein 1983] A. Weinstein, "The local structure of Poisson manifolds", *J. Differential Geom.* **18**:3 (1983), 523–557. MR 86i:58059 Zbl 0524.58011

Received May 27, 2006. Revised September 27, 2006.

PANTELIS A. DAMIANOU
DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF CYPRUS
P.O. BOX 20537
1678 NICOSIA
CYPRUS
damianou@ucy.ac.cy

HERVÉ SABOURIN
LABORATOIRE DE MATHÉMATIQUES
UMR 6086 DU CNRS
UNIVERSITÉ DE POITIERS
86962 FUTUROSCOPE CHASSENEUIL CEDEX
FRANCE
herve.sabourin@math.univ-poitiers.fr

POL VANHAECKE
LABORATOIRE DE MATHÉMATIQUES
UMR 6086 DU CNRS
UNIVERSITÉ DE POITIERS
86962 FUTUROSCOPE CHASSENEUIL CEDEX
FRANCE
pol.vanhaecke@math.univ-poitiers.fr
http://www-math.sp2mi.univ-poitiers.fr/~vanhaeck/

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

STILL ANOTHER APPROACH TO THE BRAID ORDERING

PATRICK DEHORNOY

We develop a new approach to the linear ordering of the braid group B_n, based on investigating its restriction to the set $\mathrm{Div}(\Delta_n^d)$ of all divisors of Δ_n^d in the monoid B_∞^+, that is, to positive n-braids whose normal form has length at most d. In the general case, we compute several numerical parameters attached with the finite orders $\mathrm{Div}(\Delta_n^d)$. In the case of 3 strands, we moreover give a complete description of the increasing enumeration of $\mathrm{Div}(\Delta_3^d)$. We deduce a new and especially direct construction of the ordering on B_3, and a new proof of the result that its restriction to B_3^+ is a well-ordering of ordinal type ω^ω.

This paper investigates the connection between the Garside structure of Artin's braid groups and their distinguished linear ordering, sometimes called the Dehornoy ordering. This leads to a new, alternative construction of the ordering.

Artin's braid groups B_n are endowed with several interesting combinatorial structures. One of them stems from Garside's analysis [1969] and is now known as a Garside structure [Dehornoy 2002; McCammond 2005]. It describes B_n as the group of fractions of a monoid B_n^+ with a rich divisibility theory. This theory gives a unique normal decomposition of every braid in B_n into simple braids, which are the divisors of Garside's fundamental braid Δ_n, a finite family of B_n^+ that is in one-to-one correspondence with the permutations of n objects. One obtains a natural graduation of the monoid B_n^+ by considering the family $\mathrm{Div}(\Delta_n^d)$ of all divisors of Δ_n^d, which also are the elements of B_n^+ whose normal forms have length at most d.

On the other hand, the braid groups are equipped with a distinguished linear ordering which is compatible with multiplication on the left and admits a simple combinatorial characterization [Dehornoy 1994]: a braid x is smaller than another braid y if, among all expressions of the quotient $x^{-1}y$ in the standard generators σ_i, there exists at least one expression in which the generator σ_m with maximal (or minimal) index m appears only positively, that is, σ_m occurs, but σ_m^{-1} does not. Several deep results about that ordering have been proved, for example, that

MSC2000: primary 20F36; secondary 05A05, 20F60.
Keywords: braid group, orderable group, well-ordering, normal form, fundamental braid.

its restriction to B_∞^+ is a well-ordering. A number of equivalent constructions are known [Dehornoy et al. 2002].

Although both are combinatorial, the previous structures remain mostly unconnected—and connecting them is among the most natural questions of braid combinatorics. For degree 1, that is, for simple braids, the linear ordering corresponds to a lexicographical ordering of the associated permutations [Dehornoy 1999]. But this connection does not extend to higher degrees, and almost nothing is known about the restriction of the linear ordering to positive braids of a given degree. In particular, no connection is known between the Garside normal form and the alternative normal form constructed by S. Burckel [1997; 1999; 2001] which makes comparison with respect to the linear ordering easy. For example, the Garside normal form of Δ_3^{2d} is $(\sigma_1\sigma_2\sigma_1)^{2d}$, while its Burckel normal form is $(\sigma_2\sigma_1^2\sigma_2)^d\sigma_1^{2d}$.

This paper investigates the finite linearly ordered sets $(\mathrm{Div}(\Delta_n^d), <)$. A nice way of thinking about this structure is to view the increasing enumeration of $\mathrm{Div}(\Delta_n^d)$ as a distinguished path from 1 to Δ_n^d in the Cayley graph of B_n. Completely describing this path would arguably solve optimally the rather vague task of connecting the Garside and the ordered structures of braid groups. The combinatorics of such a description seems to be extremely intricate, and it remains out of reach for the moment, but we prove partial results in this direction.

(i) In the general case, we determine some numerical parameters associated with $(\mathrm{Div}(\Delta_n^d), <)$, which in some sense measure its size. For small values of n and d, we find explicit values.

(ii) In the special case $n = 3$, we completely describe the increasing enumeration of $(\mathrm{Div}(\Delta_n^d), <)$.

Specifically, the parameters we investigate are the complexity and the heights. The complexity $c(\Delta_n^d)$ is defined as the maximal number of σ_{n-1} occurring in an expression of Δ_n^d containing no σ_{n-1}^{-1}. We connected the complexity with the termination of the handle reduction algorithm in [Dehornoy 1997], but left its determination as an open question. The r-height $h_r(\Delta_n^d)$ is defined to be the number of r-jumps in the increasing enumeration of $(\mathrm{Div}(\Delta_n^d), <)$ (augmented by 1), where the term r-jump refers to some natural filtration of the linear ordering $<$ by a sequence of partial orderings $<_r$. When r increases, the r-jumps are higher and higher, so $h_r(\Delta_n^d)$ counts how many big jumps exist in $(\mathrm{Div}(\Delta_n^d), <)$. Here, we prove that the complexity $c(\Delta_n^d)$ equals the height $h_{n-1}(\Delta_n^d)$ (Proposition 2.19), and that, for each r, the r-height $h_r(\Delta_n^d)$ is the number of divisors of Δ_n^d whose d-th factor of the normal form is right divisible by Δ_r (Proposition 3.11). Together with the combinatorial results of [Dehornoy 2007], this allows for computing the explicit values listed in Table 1, and for establishing various inductive formulas (Propositions 3.15 and 3.17, among others).

Besides the enumerative results, we also prove a general structural result that connects the ordered set $(\mathrm{Div}(\Delta_n^d), <)$ with subsets of $(\mathrm{Div}(\Delta_{n-1}^d), <)$ (Corollary 3.6). This result suggests an inductive method for directly constructing the increasing enumeration of $(\mathrm{Div}(\Delta_n^d), <)$ starting from those of $(\mathrm{Div}(\Delta_{n-1}^d), <)$ and $(\mathrm{Div}(\Delta_n^{d-1}), <)$. This approach is completed here for $n = 3$ (Proposition 4.6). In some sense, 3 strand braids are simple objects, and the result may appear as of only modest interest; however, the order on B_3^+ is a well-ordering of ordinal type ω^ω and hence not such a simple object. The interesting point is that this approach leads to a new, alternative construction of the braid ordering, with, in particular, a new and simple proof for the so-called Comparison Property at the heart of the construction (it guarantees the ordering's linearity). In this way, one obtains not just another ordering construction among many [Dehornoy et al. 2002] but, arguably, the optimal one. After the initial inductive definition is correctly stated, it makes all proofs simple and also makes explicit the connection to the Garside structure.

The paper is organized as follows. After an introductory section recalling basic properties and setting the notation, we introduce the parameters $c(\Delta_n^d)$ and $h_r(\Delta_n^d)$ in Section 2 and establish how they are connected. In Section 3, we connect in turn $h_r(\Delta_n^d)$ to the number of n-braids whose d-th factor in the normal form satisfies certain constraints, and deduce explicit values. Finally, in Section 4, we study $(\mathrm{Div}(\Delta_3^d), <)$, describe its increasing enumeration, and construct its braid ordering.

1. Background and preliminary results

Our notation is standard, and we refer to textbooks like [Birman 1974] or [Epstein et al. 1992] for basic results about braid groups. We recall that the n strand braid group B_n is defined for $n > 1$ by the presentation

$$B_n = \left\langle \sigma_1, \ldots, \sigma_{n-1} \,;\, \begin{array}{ll} \sigma_i\sigma_j = \sigma_j\sigma_i & \text{for } |i-j| \geqslant 2 \\ \sigma_i\sigma_j\sigma_i = \sigma_j\sigma_i\sigma_j & \text{for } |i-j| = 1 \end{array} \right\rangle, \tag{1-1}$$

while, for $n = 1$, we let B_1 be the trivial group. The next group B_2 is freely generated by σ_1. The elements of B_n are called n strand braids, or simply *n-braids*. We use B_∞ for the group generated by an infinite sequence of σ_i's subject to the relations of (1-1), that is, the direct limit of all B_n's with respect to the inclusion of B_n into B_{n+1}.

By definition, every n-braid x admits (infinitely many) expressions in terms of the generators σ_i, $1 \leqslant i < n$. Such an expression is called an n strand *braid word*. Two braid words w, w' representing the same braid are said to be *equivalent*; the braid represented by a braid word w is denoted $[w]$.

1A. ***Positive braids and the element Δ_n.*** We denote by B_n^+ the monoid admitting the presentation (1-1), and by B_∞^+ the union (direct limit) of all B_n^+'s. The elements of B_n^+ are called *positive* n-braids. In B_∞^+, no element except 1 is invertible, and we have a natural notion of divisibility:

Definition 1.1. For x, y in B_n^+, we say that x is a *left divisor* of y, denoted $x \preccurlyeq y$, or, equivalently, that y is a *right multiple* of x, if $y = xz$ holds for some z in B_n^+. We denote by $\mathrm{Div}(y)$ the (finite) set of all left divisors of y in B_n^+.

The monoid B_n^+ is not commutative for $n \geqslant 3$, and therefore there are distinct, but symmetric, notions of a right divisor and a left multiple; however, we shall mostly use left divisors. Note that x is a (left) divisor of y in the sense of B_n^+ if and only if it is a (left) divisor in the sense of B_∞^+, so there is no need to specify the index n.

According to Garside theory [1969], B_n^+ equipped with the left divisibility relation is a lattice: any two positive n-braids x, y admit a greatest common left divisor $\gcd(x, y)$, and a least common right multiple $\mathrm{lcm}(x, y)$. A special role is played by the lcm Δ_n of $\sigma_1, \ldots, \sigma_{n-1}$, which can be defined inductively by

$$\Delta_1 = 1, \qquad \Delta_n = \sigma_1\sigma_2\ldots\sigma_{n-1}\,\Delta_{n-1}. \tag{1-2}$$

It is well known that Δ_n^2 belongs to the center of B_n (and even generates it for $n \geqslant 3$), and that the flip automorphism ϕ_n of B_n corresponding to conjugation by Δ_n exchanges σ_i and σ_{n-i} for $1 \leqslant i \leqslant n-1$.

In B_n^+, the left and the right divisors of Δ_n coincide, and they make a finite sublattice of $(B_n^+, \preccurlyeq)$ with $n!$ elements. These braids will be called *simple*. When braid words are represented by diagrams as mentioned in Figure 1, simple braids are those positive braids that can be represented by a diagram in which any two strands cross at most once.

By mapping σ_i to the transposition $(i, i+1)$, one defines a surjective homomorphism π of B_n onto the symmetric group $\mathfrak{S}_n$. The restriction of π to simple braids is a bijection: for every permutation f of $\{1, \ldots, n\}$, there exists exactly one

Figure 1. One associates to every n strand braid word w an n strand braid diagram by stacking elementary diagrams as above. Two braid words are equivalent if and only if the associated diagrams are the projections of ambient isotopic figures in $\mathbb{R}^3$, that is, one can deform one diagram into the other without allowing the strands to cross or moving the endpoints.

simple braid x satisfying $\pi(x) = f$. It follows that the number of simple n-braids is $n!$.

Example 1.2. The set $\mathrm{Div}(\Delta_3)$ consists of six elements, namely 1, σ_1, σ_2, $\sigma_2\sigma_1$, $\sigma_1\sigma_2$, and Δ_3. In examples, we shall often use the shorter notation $\mathtt{a}$ for σ_1, $\mathtt{b}$ for σ_2, etc. Thus, the six simple 3-braids are 1, $\mathtt{a}$, $\mathtt{b}$, $\mathtt{ba}$, $\mathtt{ab}$, $\mathtt{aba}$.

1B. ***The normal form.*** For each positive n-braid x distinct from 1, the simple braid $\gcd(x, \Delta_n)$ is the maximal simple left divisor of x, and we obtain a distinguished expression $x = x_1x'$ with x_1 simple. By decomposing x' in the same way and iterating, we obtain the so-called normal expression [El-Rifai and Morton 1994; Epstein et al. 1992].

Definition 1.3. A sequence $(x_1, \ldots, x_d)$ of simple n-braids is said to be *normal* if, for each k, one has $x_k = \gcd(\Delta_n, x_k \ldots x_d)$.

Clearly, each positive braid admits a unique normal expression. It will be convenient to consider the normal expression as unbounded on the right by completing it with as many trivial factors 1 as needed. In this way, we can speak of the *d-th factor* (in the normal form) of x for each positive braid x. We say that a positive braid has *degree d* if d is the largest integer such that the d-th factor of x is not 1. We shall use the following two properties of the normal form:

Lemma 1.4 [El-Rifai and Morton 1994]. *Suppose $(x_1, \ldots, x_d)$ is sequence of simple n-braids. It is normal if and only if, for each $k < d$, each σ_i that divides x_{k+1} on the left divides x_k on the right.*

Lemma 1.5 [El-Rifai and Morton 1994]. *For x a positive braid in B_n^+, the following are equivalent*:

(i) *The braid x belongs to* $\mathrm{Div}(\Delta_n^d)$, *that is, is a (left or right) divisor of Δ_n^d*;

(ii) *The degree of x is at most d.*

Example 1.6. There are 19 divisors of Δ_3^2, which also are the 3-braids of degree at most 2. Their enumeration in normal form—in an ordering that may seem strange now, but should become familiar soon—is: 1, $\mathtt{a}$, $\mathtt{a}\cdot\mathtt{a}$, $\mathtt{b}$, $\mathtt{ba}$, $\mathtt{ba}\cdot\mathtt{a}$, $\mathtt{b}\cdot\mathtt{b}$, $\mathtt{b}\cdot\mathtt{ba}$, $\mathtt{ab}$, $\mathtt{aba}$, $\mathtt{aba}\cdot\mathtt{a}$, $\mathtt{ab}\cdot\mathtt{b}$, $\mathtt{ab}\cdot\mathtt{ba}$, $\mathtt{a}\cdot\mathtt{ab}$, $\mathtt{aba}\cdot\mathtt{b}$, $\mathtt{aba}\cdot\mathtt{ba}$, $\mathtt{ba}\cdot\mathtt{ab}$, $\mathtt{aba}\cdot\mathtt{ab}$, $\mathtt{aba}\cdot\mathtt{aba}$.

By Lemma 1.5, every divisor of Δ_n^d can be expressed as the product of at most d divisors of Δ_n, so we certainly have $\#\mathrm{Div}(\Delta_n^d) \leqslant (n!)^d$ for all n, d.

1C. ***The braid ordering.***

Definition 1.7. Let w be a nonempty braid word. We say that σ_m is the *main* generator in w if σ_m or σ_m^{-1} occurs in w, but no $\sigma_i^{\pm 1}$ with $i > m$ does. We say that w is *σ-positive* if the main generator occurs only positively in w, and similarly it is *σ-negative* if that generator occurs negatively.

A positive nonempty braid word, that is, one that contains no σ_i^{-1} at all, is σ-positive, but the inclusion is strict: for instance, $\sigma_1^{-1}\sigma_2$ is not positive, but it is σ-positive, as its main generator, namely σ_2, occurs positively (with one σ_2) but not negatively (no σ_2^{-1}).

The following two properties have received a number of independent proofs [Dehornoy et al. 2002]:

Property A. *A σ-positive braid word does not represent* 1.

Property C. *Every braid except* 1 *can be represented by a σ-positive word or by a σ-negative word.*

Building on these results, it is straightforward to order the braids:

Definition 1.8. If x, y are braids, we say that $x < y$ holds if the braid $x^{-1}y$ admits at least one σ-positive representative.

It is clear that the relation $<$ is transitive and compatible with multiplication on the left; Property A implies that $<$ has no cycle and hence is a strict partial order, and Property C then implies that it is actually a linear order.

As every nonempty positive braid word is σ-positive, $x \preccurlyeq y$ implies $x \leqslant y$ for all positive braids x, y. The converse is not true: σ_1 is not a left divisor of σ_2, but $\sigma_1 < \sigma_2$ holds because $\sigma_1^{-1}\sigma_2$ is a σ-positive word.

Example 1.9. The increasing enumeration of the set $\mathrm{Div}(\Delta_3)$ is

$$1 < \mathtt{a} < \mathtt{b} < \mathtt{ba} < \mathtt{ab} < \mathtt{aba}.$$

For instance, we have $\mathtt{ba} < \mathtt{ab}$, that is, $\sigma_2\sigma_1 < \sigma_1\sigma_2$ because the quotient, namely $\sigma_1^{-1}\sigma_2^{-1}\sigma_1\sigma_2$ (or $\mathtt{ABab}$), also admits the expression $\sigma_2\sigma_1^{-1}$, a σ-positive word. Similarly, the reader can check that the increasing enumeration of $\mathrm{Div}(\Delta_3^2)$ is the one given in Example 1.6.

Lemma 1.10. *The linear ordering $<$ extends the left divisibility ordering $\prec$.*

Proof. By definition, $1 < \sigma_i$ holds for every i. As the ordering $<$ is compatible with multiplication on the left, it follows that $x < x\sigma_i$ holds for all i, x, and, therefore, $x < xy$ holds whenever y is a nontrivial positive braid. □

Lemma 1.10 implies that 1 is always the first element of $(\mathrm{Div}(\Delta_n^d), <)$, and Δ_n^d is always its last element. A deep result by Laver [1996] shows that, although $<$ is not compatible with right multiplication in general, nevertheless $x < \sigma_i x$ always holds, that is, $<$ also extends the right divisibility ordering.

By Property C, every nontrivial braid admits at least one σ-positive or σ-negative expression. In general, such a σ-positive or σ-negative expression is not unique, but the main generator in such expressions is uniquely defined:

Lemma 1.11. *If a braid x admits a σ-positive expression, then the main generators in any two σ-positive expressions of x coincide.*

Proof. Assume that w, w' are σ-positive expressions of x, and let σ_m, $\sigma_{m'}$ be their main generators. Assume for instance $m < m'$. Then $w^{-1}w'$ is a σ-positive word, and it represents the trivial braid 1: this contradicts Property A. □

Thus, there will be no ambiguity in referring to *the* main generator of some nontrivial braid x: this means the main generator in any σ-positive (or σ-negative) expression of x.

Remark 1.12. Our definition corresponds to the order $<^{\phi}$ of [Dehornoy et al. 2002]. It differs from the one most used in the literature in that the definition of a σ-position refers to the maximal index rather than the minimal one. Switching from one definition to the other amounts to conjugating by Δ_n, that is, to applying the flip automorphism. The results are entirely similar for both versions. However, it is much more convenient to consider the "max" choice here, because it guarantees that B_n^+ is an initial segment of B_{n+1}^+. Using the "min" convention would make the statements in the following sections less natural.

2. Measuring the ordered sets $(\mathrm{Div}(\Delta_n^d), <)$

To investigate the finite ordered sets $(\mathrm{Div}(\Delta_n^d), <)$, and, more generally, the sets $(\mathrm{Div}(z), <)$ for positive braids z, we shall define numerical parameters that reflect their size. The first parameter involves the length of the σ-positive words that are, in a natural sense defined below, drawn in the Cayley graph of Δ_n^d. It will be called the *complexity* of Δ_n^d, because it is directly connected with the complexity analysis of the handle reduction algorithm of [Dehornoy 1997]. The other parameters involve a filtration of the linear ordering by the σ_i's, and they will be called the *heights* of Δ_n^d because they count the jumps of a given height in $(\mathrm{Div}(\Delta_n^d), <)$.

2A. ***Sigma-positive paths in the Cayley graph.*** The first parameter we attach to $(\mathrm{Div}(z), <)$ involves the σ-positive paths in the Cayley graph of z.

We recall that the Cayley graph of the group B_n with respect to the standard generators σ_i is a labeled graph: it has the vertex set B_n and is such that there exists a σ_i-labeled edge from x to y if and only if $y = x\sigma_i$. The Cayley graph of the monoid B_n^+ is obtained by restricting the vertices to B_n^+. Note that the Cayley graph of B_n (and a fortiori of B_n^+) can be seen as a subgraph of the Cayley graph of B_∞.

Definition 2.1. (See Figure 2.) For z a positive braid, we denote by $\Gamma(z)$ the subgraph of the Cayley graph of B_∞ obtained by restricting the vertices to $\mathrm{Div}(z)$ and removing the edges do not connect two vertices in $\mathrm{Div}(z)$.

Because every element of B_n^+ is a left divisor of Δ_n^d for sufficiently large d, the Cayley graph of B_n^+ is the union over all d of the graphs $\Gamma(\Delta_n^d)$.

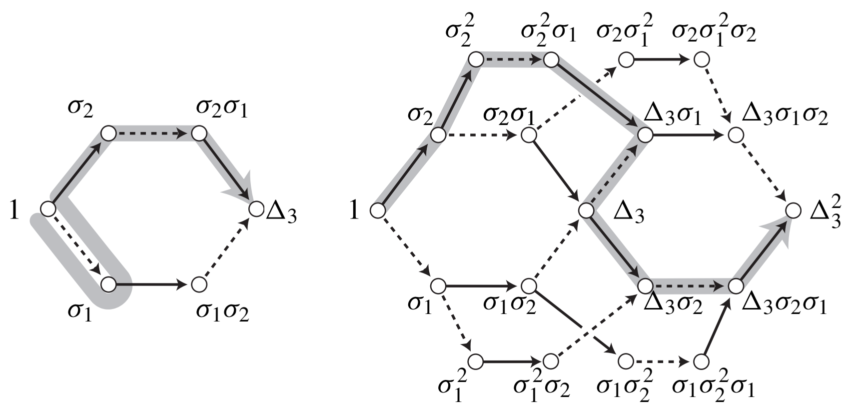


Figure 2. The graphs of $\Gamma(\Delta_3)$ and $\Gamma(\Delta_3^2)$; the dotted edges represent σ_1, the plain ones σ_2; observe that the graph of Δ_3^2 is not planar; in grey: two σ-positive words traced in the graphs, namely `aAbab` and `bbabAbab` (see Lemma 2.3).

A path in the Cayley graph can be specified by its initial vertex and the listed labels of its successive edges, that is, by a braid word. For each $i < n$ and each x in B_n, there is in B_n's Cayley graph exactly one σ_i-labeled edge leading into x and exactly one other going out of it. Hence, in the complete Cayley graph of B_n, for each initial vertex x and each n-braid word w, there is always one path labeled w starting from x. When we restrict to some fragment Γ, this need not be the case, but we do have an unambiguous notion of w being drawn in Γ from x. Formally:

Definition 2.2. If Γ is a subgraph of the Cayley graph of B_∞, and x is a vertex in Γ, we say that a braid word w is *drawn* from x in Γ if, for every prefix $v\sigma_i$ (*resp.* $v\sigma_i^{-1}$) of w, there exists a σ_i-labeled edge starting (*resp.* finishing) at $x\,[v]$ in Γ.

For instance, we can check on Figure 2 that the word σ_1^2 is drawn from σ_2 in $\Gamma(\Delta_3^2)$, but not in $\Gamma(\Delta_3)$. In algebraic terms,

Lemma 2.3. *Assume that z is a positive braid, and w is a braid word. Then w is drawn from x in $\Gamma(z)$ if and only if $x[v] \preccurlyeq z$ holds for each prefix v of w.*

Proof. The condition is sufficient. Indeed, assume it is satisfied by w, and $v\sigma_i$ is a prefix of w. Then, by hypothesis, $x[v]$ and $x[v]\sigma_i$ are left divisors of z. Hence are vertices in $\Gamma(z)$, and, therefore, there is a σ_i-labeled edge between $x[v]$ and $x[v]\sigma_i$ in $\Gamma(z)$. The argument is similar for a prefix of the form $v\sigma_i^{-1}$. Using induction on the length of w, we deduce that w is drawn from x in $\Gamma(z)$.

Conversely, if there is a w-labeled path from x in $\Gamma(z)$, then, for each prefix v of w, the braid $x[v]$ represents some vertex in $\Gamma(z)$. Hence it's a left divisor of z. □

For z a positive braid, we shall investigate the σ-positive words drawn in the graph $\Gamma(z)$. It is clear that, even if $\mathrm{Div}(z)$ is a finite set, arbitrary long words are drawn in $\Gamma(z)$ whenever the latter contains at least 2 vertices, that is, z is not 1. The example of Figure 2 shows that restricting to σ-positive words does not change the result: for instance, for each k, the word $(\sigma_1\sigma_1^{-1})^k\sigma_2\sigma_1\sigma_2$ is a σ-positive expression of Δ_3, and it is drawn in $\Gamma(\Delta_3)$. So we cannot hope for any finite upper bound on the length of the σ-positive words drawn in $\Gamma(z)$ in general. The situation changes if we concentrate on the main generators, that is, we forget about the generators with nonmaximal index.

Lemma 2.4. *Assume that Γ is subgraph of the Cayley graph of B_∞, and w is a σ-positive word drawn in $\Gamma(z)$. Then the number of occurrences of the main generator in w is at most the number of nonterminal vertices in Γ.*

Proof. Assume that w is drawn from x in Γ. Let σ_m be the main generator in w. As there is at most one σ_m-labeled edge starting from each vertex of Γ, it suffices to show that the number of σ_m's in w is bounded above by the number of σ_m-edges in Γ. Hence, it suffices to show that the path γ associated with w cannot cross the same σ_m-edge twice. Now assume that some σ_m-edge starts from the vertex y, and that γ crosses this edge twice. This means that γ contains a loop from y to y. Let v be the subword of w labeling that loop. By construction, v begins with σ_m, it contains no σ_m^{-1} and no $\sigma_i^{\pm 1}$ with $i > m$ as it is a subword of w, and it represents the braid 1 as it labels a loop in the Cayley graph of B_∞: this means that v is a σ-positive word representing 1, which contradicts Property A. □

Lemma 2.4 applies in particular to every graph $\Gamma(z)$ in which z is a positive braid. We can introduce our first parameter measuring the size of the ordered set $(\mathrm{Div}(z), <)$:

Definition 2.5. (See Figure 2.) When z is a positive braid with main generator σ_m, the *complexity* $c(z)$ of z is defined to be the maximal number of σ_m's in a σ-positive word drawn in $\Gamma(z)$.

Example 2.6. The word $\sigma_2\sigma_1\sigma_2$ is a σ-positive word drawn from 1 in $\Gamma(\Delta_3)$, and it contains two σ_2's. Hence we have $c(\Delta_3) \geqslant 2$. Actually, it is not hard to obtain the exact value $c(\Delta_3) = 2$. Indeed, if a σ-positive path γ contains the two σ_2-edges starting from 1 and $\sigma_1\sigma_2$, it cannot come back to σ_2 without crossing the third σ_2-edge; and if γ contains the σ_2-edge that starts from σ_1, it can never come back to 1 or to $\sigma_2\sigma_1$ and therefore contains at most one σ_2-edge. As we have $\Delta_3^d = (\sigma_2\sigma_1\sigma_2)^d$, we deduce $c(\Delta_3^d) \geqslant 2d$ for every d. This value is certainly not optimal: Figure 2 contains five σ_2's, proving $c(\Delta_3^2) \geqslant 5$. The exact value here is 6, and, more generally, we have $c(\Delta_n^d) = 2^{d+1} - 2$, as will be seen in Section 3.

Remark 2.7. Restricting to σ-positive words drawn in $\Gamma(z)$ is essential: for instance, for each k, we have

$$\Delta_3 = \sigma_2^{k+1}\sigma_1\sigma_2\sigma_1^{-k}, \tag{2-1}$$

a σ-positive word containing $k+2$ letters σ_2. Now, for $k \geqslant 1$, the word involved in (2-1) is not drawn in $\Gamma(\Delta_3^1)$, because its prefix σ_2^2 is not. Thus the parameter $c(z)$ does involve the left divisors of z.

Directly applying Lemma 2.4 gives:

Proposition 2.8. *Every positive braid has a finite complexity*; *more precisely*, *for z of length ℓ in B_n^+ with $n \geqslant 3$, we have $c(z) \leqslant (n-1)^\ell$.*

Proof. The number of nonterminal vertices in $\Gamma(z)$, that is, the number of proper left divisors of z, is at most $1 + (n-1) + (n-1)^2 + \cdots + (n-1)^{\ell-1}$. □

As the length of any positive expression of Δ_n is $n(n-1)/2$, we obtain in particular for all n, d

$$c(\Delta_n^d) \leqslant (n-1)^{dn(n-1)/2}. \tag{2-2}$$

Before going further, we observe that, in defining the complexity of z, we can restrict to decompositions of z, that is, instead of considering paths starting and finishing at arbitrary vertices, we can restrict to paths going from 1 to z:

Lemma 2.9. *Assume that z is a positive braid with main generator σ_m. Then $c(z)$ is the maximal number of σ_m's in any σ-positive decomposition of z drawn in $\Gamma(z)$.*

Proof. Let $c'(z)$ be the number involved in the above statement. Clearly we have $c'(z) \leqslant c(z)$. Conversely, assume that w is drawn in $\Gamma(z)$ from x, and that the w-labeled path starting at x finishes at y. Let u be a positive expression of x, and v be a positive expression of $y^{-1}z$. The latter exists as, by hypothesis, y is a left divisor of z. Then uwv is a σ-positive decomposition of z drawn in $\Gamma(z)$. Hence we have $c'(z) \geqslant c(z)$. □

Remark 2.10. We call Property A* the statement that all numbers $c(\Delta_n^d)$ are finite. Above, we derived Property A* from Property A. The two properties are actually equivalent, that is, we can also deduce Property A from Property A*. For that, assume that some σ-positive braid word w represents 1. The word w may involve negative letters. We must find a vertex x that begins a path labeled w in some $\Gamma(\Delta_n^d)$. Let σ_m be the main generator in w. The word w has finitely many prefixes, say $w_0, \ldots, w_\ell$. By Garside theory, each word w_i is equivalent to one the form

$u_i^{-1}v_i$, with u_i, v_i positive. Let x be the least common left multiple of the positive braids

$$[u_0], \ldots, [u_\ell].$$

For each i, the braid $x[w_i]$ is positive. Moreover, there exist n and d such that

$$x[w_0], \ldots, x[w_\ell]$$

are all divisors of Δ_n^d. Thus the word w is drawn from x in $\Gamma(\Delta_n^d)$, and the associated path is a loop around x. It follows that w^k is drawn in $\Gamma(\Delta_n^d)$ from x for each k. By construction, w^k contains at least k generators σ_m. Hence $c(\Delta_n^d)$ cannot be finite.

2B. ***Connection with handle reduction.*** Handle reduction [Dehornoy 1997] is an algorithmic solution to the word problem of braids that relies on the braid ordering. It is the most efficient method today. The method converges, and the argument in [Dehornoy 1997] shows the complexity upper bound to be exponential in the input word length, an estimate seemingly very far from sharp.

Each step of handle reduction involves a specific generator σ_i, and, for an induction, the point is to obtain an upper bound on the reduction steps involving the main generator. The latter will naturally be called the *main* reduction steps. The connection between handle reduction and the complexity as defined above relies on the following technical result:

Lemma 2.11 [Dehornoy 1997]. *Assume that z is a positive braid with main generator σ_m and that w is drawn in $\Gamma(z)$. Then, for each sequence of handle reductions from w — that is, each sequence $\vec{w}$ with $w_0 = w$ such that w_k is obtained by reducing one handle from w_{k-1} for each k — there exists a witness-word u that is σ-positive, drawn in* $\mathrm{Div}(z)$, *and such that the number of σ_m's in u is the number of main reductions in $\vec{w}$.*

It follows that the number of main reduction steps in any sequence of handle reductions starting with a word drawn in $\Gamma(z)$ is bounded above by $c(z)$. In particular, if we start with an n strand braid word w of length ℓ, then it is easy to show that w is drawn in $\Gamma(\Delta_n^\ell)$, and, applying the upper bound of Equation (2-2), we deduce the upper bound on the number of possible main reductions from w, and it is exponential in ℓ.

A natural way to improve this coarse upper bound would be to determine $c(\Delta_n^d)$ more precisely. This will be done in Section 3 below. However, the explicit formulas show that, for $n \geqslant 3$, the growth in d really is exponential, thus dashing any hopes of proving a polynomial upper bound for the number of reduction steps by this approach.

2C. ***A filtration of the braid ordering.*** We now introduce new numerical parameters for the ordered sets $(\mathrm{Div}(z), <)$. These numbers connect with a natural filtration of the ordering $<$, using an increasing sequence of partial orderings.

By Lemma 1.11, the index of the main generator of a nontrivial braid is well defined. We can use this index to measure the height of the jump between two braids x, y satisfying $x < y$:

Definition 2.12. For x, y in B_∞ and $r \geqslant 1$, we say that $x <_r y$ holds or, equivalently, that (x, y) is an *r-jump*, if $x^{-1}y$ admits a σ-positive expression in which the main generator is σ_m with $m \geqslant r$.

Lemma 2.13. *For each $r \geqslant 1$, the relation $<_r$ is a strict partial order that refines $<$; the relation $<_1$ coincides with $<$, and $r \leqslant q$ implies that $<_q$ refines $<_r$.*

Proof. That $<_r$ is transitive follows because the concatenation of a σ-positive word with main generator σ_m and a σ-positive word with main generator $\sigma_{m'}$ is a σ-positive word with main generator $\sigma_{\max(m,m')}$. □

In the sequel, we consider the $<_r$-chains included in $\mathrm{Div}(z)$, and their length:

Definition 2.14. For z a positive braid and $r \geqslant 1$, we define the *r-height* $h_r(z)$ of z to be the maximal length of a $<_r$-chain included in $\mathrm{Div}(z)$.

Before giving examples, we observe the connection between $h_r(z)$ and the increasing enumeration of the set $\mathrm{Div}(z)$:

Lemma 2.15. *Let z be a positive braid and $r \geqslant 1$. Then $h_r(z) - 1$ is the number of r-jumps in the increasing enumeration of* $(\mathrm{Div}(z), <)$.

Proof. If the number of r-jumps in the increasing enumeration of $\mathrm{Div}(z)$ is $N_r - 1$, we can extract from $\mathrm{Div}(z)$ a $<_r$-chain of length N_r. Conversely, assume that $(y_0, \ldots y_{N_r})$ is a $<_r$-chain in $\mathrm{Div}(z)$. Let $z_0 < \ldots < z_N$ be the increasing enumeration of $\mathrm{Div}(z)$. As $<_r$ refines $<$, there exists an increasing function f of $\{0, \ldots, N_r\}$ into $\{0, \ldots, N\}$ such that $y_i = z_{f(i)}$ holds for every i. Now the hypothesis $z_{f(i)} <_r z_{f(i+1)}$ implies that there exists at least one r-jump between $z_{f(i)}$ and $z_{f(i+1)}$. Indeed, by Lemma 1.11, it is impossible that a concatenation of m-jumps with $m < r$ results in a r-jump. So the number of r-jumps in $(z_0, \ldots, z_N)$ is at least N_r. □

In other words, to determine $h_r(z)$, there is no need to consider arbitrary chains: it is enough to consider the maximal chain obtained by enumerating $\mathrm{Div}(z)$ exhaustively.

Example 2.16. Refining the increasing enumeration of $\mathrm{Div}(\Delta_3)$ of Example 1.9 by indicating for each step the height of the corresponding jump, we obtain:

$$1 <_1 \mathtt{a} <_2 \mathtt{b} <_1 \mathtt{ba} <_2 \mathtt{ab} <_1 \Delta_3, \tag{2-3}$$

where we recall $\mathtt{a}, \mathtt{b}, \ldots$ stand for $\sigma_1, \sigma_2, \ldots$. For instance, $(\mathtt{ba}, \mathtt{ab})$ is a 2-jump, because we have in $(\mathtt{ba})^{-1}(\mathtt{ab}) = \mathtt{ABab} = \mathtt{AabA} = \mathtt{bA}$ a σ-positive decomposition with main generator σ_2. The number of 1-jumps in (2-3), that is, the number of symbols $<_r$ with $r \geqslant 1$, is 5, while the number of 2-jumps is 2, so, by Lemma 2.15, we deduce $h_1(\Delta_3) = 6$ and $h_2(\Delta_3) = 3$. Similarly, we obtain for Δ_3^2

$$\mathtt{1} <_1 \mathtt{a} <_1 \mathtt{aa} <_2 \mathtt{b} <_1 \mathtt{ba} <_1 \mathtt{baa} <_2 \mathtt{bb} <_1 \mathtt{bba} <_2 \mathtt{ab} <_1 \mathtt{aba} <_1 \mathtt{abaa} <_2 \mathtt{abb}$$
$$<_1 \mathtt{abba} <_2 \mathtt{aab} <_1 \mathtt{aaba} <_1 \mathtt{aabaa} <_2 \mathtt{baab} <_1 \mathtt{baaba} <_1 \mathtt{baabaa},$$

leading to $h_1(\Delta_3^2) = 19$ and $h_2(\Delta_3^2) = 7$.

Proposition 2.17. (i) *For every braid z in B_n^+, we have*

$$h_1(z) = \#\mathrm{Div}(z) \geqslant h_2(z) \geqslant \cdots \geqslant h_n(z) = 1.$$

(ii) *For all positive braids z, z' and $r \geqslant 1$, we have*

$$h_r(zz') \geqslant h_r(z) + h_r(z'). \tag{2-4}$$

Proof. (i) A $<_1$-chain is simply a $<$-chain. Hence every subset of $\mathrm{Div}(z)$ gives such a chain. So the maximal $<_1$-chain in $\mathrm{Div}(z)$ is $\mathrm{Div}(z)$ itself, and $h_1(z)$ is the cardinality of $\mathrm{Div}(z)$.

On the other hand, no $<_n$-chain in B_n^+ has length more than 1, as the main generator of a σ-positive n-strand braid word cannot be σ_n or any generator above it. Thus $h_n(z)$ is 1.

Then, for $q \leqslant r$, every $<_r$-chain is a $<_q$-chain, which implies $h_r(z) \geqslant h_q(z)$.

Point (ii) is obvious, as the concatenation of two $<_r$-chains is a $<_r$-chain. □

From (2-4) we deduce $h_r(z^d) \geqslant d \cdot h_r(z)$ for all r, z. By Lemma 1.5, every divisor of Δ_n^d can be decomposed as the product of at most d divisors of Δ_n. There are $n!$ such divisors, so we obtain the (coarse) bounds

$$d \cdot h_r(\Delta_n) \leqslant h_r(\Delta_n^d) \leqslant (n!)^d,$$

for all r, n, d. Better estimates will be given below.

Remark 2.18. Instead of restricting to subsets of B_∞ of the form $\mathrm{Div}(z)$, we can define the complexity and the r-height for every (finite) set of braids X. Most of the general results extend, but, when X is not closed under left division, nothing can be said about the number of σ_r's involved in an r-jump. Considering such an extension is not useful here.

2D. ***Connection with the complexity.*** We shall now connect the complexity $c(z)$ with the numbers $h_r(z)$ just defined. The result is simple:

Proposition 2.19. *For z a positive braid with main generator σ_m, we have*

$$c(z) = h_m(z) - 1.$$

In particular, for $n \geqslant 2$ and $d \geqslant 0$, we have

$$c(\Delta_n^d) = h_{n-1}(\Delta_n^d) - 1.$$

One inequality is easy:

Lemma 2.20. *For z a positive braid with main generator σ_m, we have $c(z) \leqslant h_m(z) - 1$.*

Proof. The argument is reminiscent of the one used for Lemma 2.15 but requires a little more care. Assume that w is a σ-positive word drawn in $\Gamma(z)$ from x containing N_m occurrences of σ_m. By Lemma 2.9, we can assume $x = 1$ without loss of generality. Let $z_0 < z_1 < \ldots < z_N$ be the increasing enumeration of $\mathrm{Div}(z)$. By definition, all prefixes of w represent divisors of z, so, letting ℓ be the length of w, there exists a mapping $f : \{0, \ldots, \ell\} \to \{0, \ldots, N\}$ such that, for each k, the length k prefix of w represents $z_{f(k)}$. By construction, we have $f(0) = 0$ and $f(\ell) = N$.

The difference from Lemma 2.15 is that f need not be increasing. Now, let $p_1, \ldots, p_{N_m}$ be the N_m positions in w where the generator σ_m occurs, completed with $p_0 = 0$. Then, in the prefix of w of length p_1, that is, in the subword of w corresponding to positions from $p_0 + 1$ to p_1, there is one σ_m, plus letters $\sigma_i^{\pm 1}$ with $i < m$ (Figure 3). This subword is therefore σ-positive. Hence we must have $z_{f(p_0)} < z_{f(p_1)}$, which requires $f(p_0) < f(p_1)$. Moreover, the quotient $z_{f(p_0)}^{-1} z_{f(p_1)}$ is a braid that admits at least one σ-positive expression containing σ_m, and hence $z_{f(p_0)} <_m z_{f(p_1)}$. Now the same is true between $f(p_1)$ and $f(p_2)$, etc. Hence the number of m-jumps in the increasing enumeration of $\mathrm{Div}(z)$ is at least N_m, that is, we have $h_m(z) \geqslant N_m + 1$. $\square$

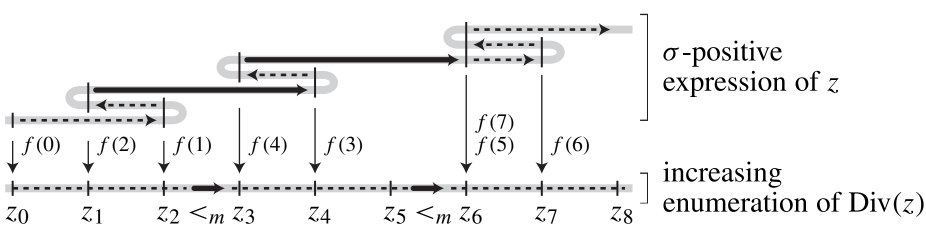


Figure 3. Proof of Lemma 2.20. The main generator σ_m corresponds to the bold arrow: the function f need not be increasing, but the projection of a bold arrow upstairs must include at least one bold arrow downstairs, that is, at least one m-jump.

It remains to prove the second inequality in Proposition 2.19, that is, to prove that, if z is a positive n-braid satisfying $h_m(z) = N + 1$, then z admits a σ-positive expression containing N generators σ_m. The problem is as follows: if z is a positive braid and x, y are left divisors of z satisfying $x < y$, then, by definition, the quotient $x^{-1}y$ admits some σ-positive expression w, but nothing *a priori* guarantees that w be drawn in $\Gamma(z)$. In other words, we might have $x < y$ but no σ-positive witness for this inequality inside $\mathrm{Div}(z)$. It turns out this cannot happen, but the proof requires a rather delicate argument.

Proposition 2.21. *Let z be a positive braid. Then, for all x, y in* $\mathrm{Div}(z)$, *the following are equivalent*:

(i) *The relation $x < y$ holds, that is, there exists a σ-positive path from x to y in the Cayley graph of B_∞*;

(ii) *There exists a σ-positive path from x to y in the Cayley graph of B_n*;

(iii) *There exists a σ-positive path from x to y in $\Gamma(z)$.*

Proof. Clearly (iii) implies (ii), which in turn implies (i). We shall prove that (i) implies (iii) — and thus reprove that (i) implies (ii), which was first proved in [Larue 1994] — by using the handle reduction method of [Dehornoy 1997; Dehornoy et al. 2002]. The problem is to prove that, among all σ-positive paths connecting x to y in the Cayley graph of B_∞, at least one is drawn in $\Gamma(z)$.

Now, let u, v be positive words representing x and y. Then the word $u^{-1}v$ represents $x^{-1}y$, and, by hypothesis, it is drawn in $\Gamma(z)$ from x. Handle reduction transforms a braid word into equivalent words and eventually produces a σ-positive word if it exists. It is proved in [Dehornoy 1997] that, for every n strand braid word w, there exists a finite fragment Γ_w of the Cayley graph of B_n^+ and a vertex x_w of Γ_w such that w and all words obtained from w by handle reduction are drawn from x_w in Γ_w. Moreover, when w has the form $u^{-1}v$ with u, v positive, then all vertices in Γ_w are the left divisors of the least common right multiple of the braids represented by u and v, here x and y, while x_w is the braid represented by u, that is, x. As x and y are divisors of z, so is their least common right multiple, and the graph Γ_w is included in $\Gamma(z)$. It follows that every word obtained from $u^{-1}v$ using handle reduction is drawn from x in $\Gamma(z)$. The termination of handle reduction guarantees that, among these words, at least one is σ-positive, so (iii) follows. □

A direct application of Proposition 2.21 is the existence of σ-positive quotient sequences drawn in the Cayley graph. The definition is as follows:

Definition 2.22. Assume that z is a positive braid and X is a subset of $\mathrm{Div}(z)$. Let $x_0 < \ldots < x_N$ be the increasing enumeration of X. We say that a sequence of words $\vec{w} = (w_1, \ldots, w_N)$ is a *quotient sequence* for X if, for each k, the word w_k is an expression of $x_{k-1}^{-1}x_k$ for each k. We say that $\vec{w}$ is *σ-positive* if every entry in

$\vec{w}$ is σ-positive, and that $\vec{w}$ is *drawn in* $\Gamma(z)$ (from x_0) if w_k is drawn from x_{k-1} in $\Gamma(z)$ for each k.

Corollary 2.23. *Assume that z is a positive braid. Then every subset of* $\mathrm{Div}(z)$ *admits a σ-positive quotient sequence drawn in* $\Gamma(z)$.

Example 2.24. (Figure 4) By computing the successive quotients in the increasing enumeration of $\mathrm{Div}(\Delta_3^2)$ given in Example 1.9, we easily find that

$$(\mathtt{a}, \mathtt{a}, \mathtt{AAb}, \mathtt{a}, \mathtt{a}, \mathtt{AAb}, \mathtt{a}, \mathtt{AAb}, \mathtt{a}, \mathtt{a}, \mathtt{bAA}, \mathtt{a}, \mathtt{bAA}, \mathtt{a}, \mathtt{a}, \mathtt{bAA}, \mathtt{a}, \mathtt{a})$$

is a σ-positive quotient sequence for $\mathrm{Div}(\Delta_3^2)$ drawn in $\Gamma(\Delta_3^2)$. This sequence turns out to be the unique sequence with the above properties, but this uniqueness is specific to the case of 3-braids (see Figure 8 below).

We can now easily complete the proof of Proposition 2.19:

Proof of Proposition 2.19. Let $(z_0, \ldots, z_N)$ be the $<$-increasing enumeration of $\mathrm{Div}(z)$. By Corollary 2.23, there exists a σ-positive quotient sequence $\vec{w}$ for $\mathrm{Div}(z)$ that is drawn in $\Gamma(z)$. Let $w = w_1 \ldots w_N$. By construction, w is a σ-positive word drawn in $\Gamma(z)$, and the number of occurrences of the main generator σ_m in w is (at least) the number of m-jumps in $(z_0, \ldots, z_N)$. So we have $c(z) \geqslant h_m(z) - 1$. Invoking Lemma 2.20 completes the proof. □

Remark 2.25. Assume that $\vec{w}$ is a σ-positive quotient sequence for $\mathrm{Div}(z)$, and σ_m is the main generator occurring in $\vec{w}$. Then each word w_i contains zero or one letter σ_m. Indeed, if w_i contained two σ_m's or more, then the vertex reached after the first σ_m ought to lie in the open $<$-interval determined by two successive entries of $\vec{z}$, and the latter is empty by construction since all elements of $\mathrm{Div}(z)$ occur in $\vec{z}$.

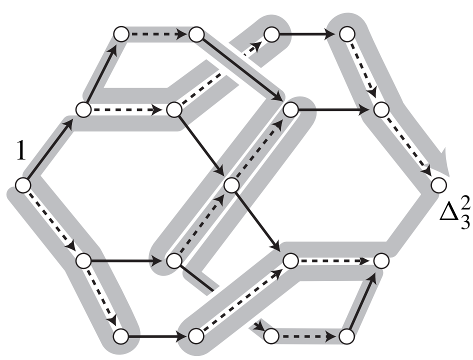


Figure 4. The increasing enumeration of the divisors of Δ_3^2, and a σ-positive quotient sequence drawn in $\Gamma(\Delta_3^2)$: the associated path visits every vertex, and is labeled aaAAbaaAAbabAAa aAAbabAAaabAAaa; it crosses 6 σ_2-edges (and no σ_2^{-1}).

3. A decomposition result for (Div(z), <)

In this section, we establish a structural result describing $(\mathrm{Div}(\Delta_n^d), <)$ as the concatenation of $c(\Delta_n^d) + 1$ intervals isomorphic to subsets of $(\mathrm{Div}(\Delta_{n-1}^d), <)$. We deduce an explicit formula connecting $h_r(\Delta_n^d)$ with the number of braids in $\mathrm{Div}(\Delta_n^d)$ whose d-th factor is right divisible by Δ_r, which in turn enables us to finish computing $c(\Delta_n^d)$ and $h_r(\Delta_n^d)$ for small values of r, n and d.

3A. ***B_r-classes.*** To analyze the linearly ordered sets $(\mathrm{Div}(\Delta_n^d), <)$, and, more generally, $(\mathrm{Div}(z), <)$ for z a positive braid, we introduce convenient partitions. As B_r is a group for each r, it is clear that the relation $x^{-1}y \in B_r$ defines an equivalence relation on (positive) braids, so we may put:

Definition 3.1. For $r \geqslant 1$ and x, y in B_∞^+, we say that x and y are *B_r-equivalent* if $x^{-1}y$ belongs to B_r.

By construction, B_r-equivalence is compatible with multiplication on the left. In the sequel, we consider the restriction of B_r-equivalence to finite subsets of B_∞^+ of the form $\mathrm{Div}(z)$, that is, we use B_r-equivalence to partition $\mathrm{Div}(z)$ into subsets, naturally called B_r-classes.

Example 3.2. As B_1 is trivial, B_1-equivalence is equality, and so, therefore, the B_1-classes are singletons. On the other hand, any two elements of B_n are B_r-equivalent for each $r \geqslant n$, so, for z in B_n^+, there is only one B_r-class for $r \geqslant n$, and the only interesting relations arise for $1 < r < n$. For instance, $\mathrm{Div}(\Delta_3)$ contains three B_2-classes, while $\mathrm{Div}(\Delta_3^2)$ contains seven of them (Figure 5).

Saying that there is an r-jump between two braids x and y means that $x^{-1}y$ is σ-positive and does not belong to B_r, so, for $x < y$, we have the equivalence

$$\bigl(x, y \text{ are not } B_r\text{-equivalent}\bigr) \Longleftrightarrow \begin{pmatrix}\text{there is a } r\text{-jump between} \\ \text{between } x \text{ and } y\end{pmatrix}. \tag{3-1}$$

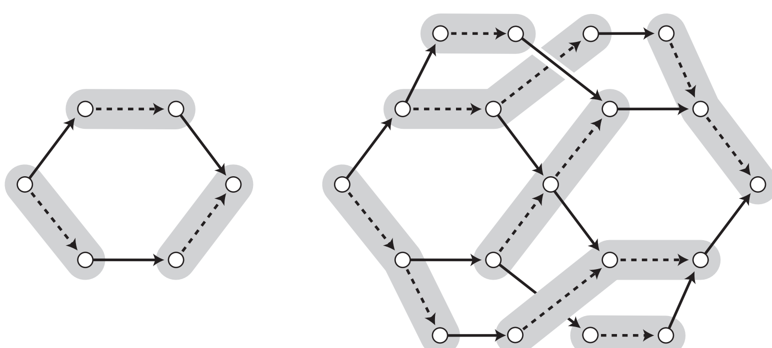


Figure 5. The B_2-classes in $\mathrm{Div}(\Delta_3)$ and $\mathrm{Div}(\Delta_3^2)$.

Lemma 3.3. *Assume that z is a positive braid. Then, each B_r-class in* $\mathrm{Div}(z)$ *is an interval for $<$ and there is an r-jump between each B_r-class and the next one.*

Proof. Assume $x < y \in \mathrm{Div}(z)$. By (3-1), if x and y are not B_r-equivalent, there is an r-jump between x and y and hence also between x and any element of $\mathrm{Div}(z)$ above y. Thus no such element may be B_r-equivalent to x. This implies that each B_r-class is an $<$-interval. □

Corollary 3.4. *For each $r \geqslant 1$, the number of B_r-classes in* $\mathrm{Div}(z)$ *is $h_r(z)$.*

Proof. By (3-1), there is no r-jump between two elements of the same B_r-class, and there is one between two elements not in the same B_r-class. Thus the number of B_r-classes is the number of r-jumps in the $<$-increasing enumeration of $\mathrm{Div}(z)$ augmented by 1. Hence, by Lemma 2.15, it is $h_r(z)$. □

With B_r-equivalence, we can partition $(\mathrm{Div}(z), <)$ into finitely many subintervals. The interest of this partition is that we can describe B_r-classes rather precisely and, typically, connect them with subsets of B_r. In particular, this will allow for connecting the ordered sets $(\mathrm{Div}(\Delta_n^d), <)$ with the sets $(\mathrm{Div}(\Delta_{n-1}^d), <)$.

Proposition 3.5 (Figure 6). *Assume $z \in B_\infty^+$ and $r \geqslant 1$. Let C be a B_r-class in* $\mathrm{Div}(z)$, *and let a, b be its $<$-extremal elements. Then a divides every element of C on the left, and the left translation by a establishes an isomorphism between* $(\mathrm{Div}(a^{-1}b), \preccurlyeq, <)$ *and $(C, \preccurlyeq, <)$. In particular, $(C, \preccurlyeq)$ is a lattice.*

Proof. By Lemma 3.3, C is the $<$-interval determined by a and b, that is, we have

$$C = \{x \in \mathrm{Div}(z);\ a < x < b\}.$$

We know that $\mathrm{Div}(z)$ is a lattice with respect to left divisibility: any two elements x, y of $\mathrm{Div}(z)$ admit a greatest common left divisor, here denoted $\gcd(x, y)$, and a least common right multiple, denoted $\mathrm{lcm}(x, y)$. Firstly, we claim that C is a lattice with respect to left divisibility, that is, the left gcd and the right lcm of two elements of C lie in C. So assume $x, y \in C$. Let x_0, y_0 be defined by $x = \gcd(x, y)x_0$ and $y = \gcd(x, y)y_0$. The hypothesis that $x^{-1}y$ belongs to B_r implies that there exist x_1, y_1 in B_r^+ satisfying $x^{-1}y = x_1^{-1}y_1$. By definition of the gcd, there must exist a positive braid z_1 satisfying $x_1 = z_1x_0$ and $y_1 = z_1y_0$. Because z_1 is positive, $x_1 \in B_r^+$ implies $x_0 \in B_r^+$, and hence $\gcd(x, y) \in C$. As for the lcm, the conjunction of $x = \gcd(x, y)x_0$ and $y = \gcd(x, y)y_0$ implies

$$\mathrm{lcm}(x, y) = \gcd(x, y)\,\mathrm{lcm}(x_0, y_0).$$

As $x_0, y_0 \in B_r^+$ implies $\mathrm{lcm}(x_0, y_0) \in B_r^+$, we deduce $\mathrm{lcm}(x, y) \in C$.

As C is finite, it follows that C admits a global gcd. Because the linear ordering $\leqslant$ extends the partial divisibility ordering $\preccurlyeq$, this global gcd must be the $<$-minimum a of C. Symmetrically, C admits a global lcm, which must be the

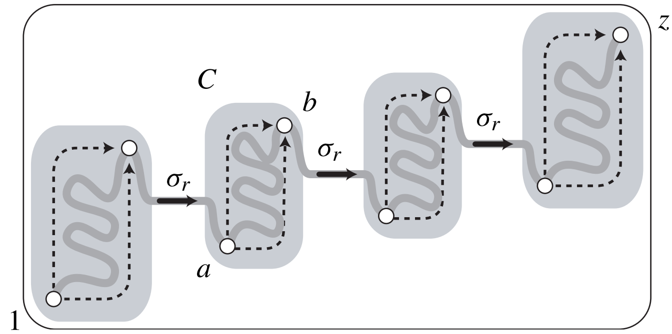


Figure 6. Decomposition of $(\mathrm{Div}(z), <)$ into B_r-classes: each class C is a lattice with respect to divisibility; the increasing enumeration of $\mathrm{Div}(z)$ exhausts the first class, then jumps to the next one by an r-jump, etc. The number of classes is $h_r(z)$.

$<$-maximum b. So, at this point, we know that a is a left divisor of every element in C, and b is a right multiple of each such element, that is, we have

$$C \subseteq \{x \in B_\infty^+ ; a \preccurlyeq x \preccurlyeq b\}. \tag{3-2}$$

Moreover, $a \preccurlyeq x \preccurlyeq b$ implies $a \leqslant x \leqslant b$. Hence $x \in C$, and so the inclusion in (3-2) is an equality.

Now, put $F(x) = ax$ for x in $\mathrm{Div}(a^{-1}b)$. As B_∞^+ is left cancellative, F is injective. Moreover, for x a positive braid, $x \preccurlyeq a^{-1}b$ is equivalent to $ax \preccurlyeq b$, so the image of F is $\{x \in B_\infty^+ ; a \preccurlyeq x \preccurlyeq b\} = C$. Finally, by construction, F preserves both $\preccurlyeq$ and $<$. □

For $r = 1$, each B_r-class is a singleton, and Proposition 3.5 says nothing; similarly, if the main generator of z is σ_m, there is only one B_r-class for $r > m$, and we gain no information. But, for $1 < r \leqslant m$, and specially for $r = m$, Proposition 3.5 states that the chain $\mathrm{Div}(z)$ is obtained by concatenating $h_r(z)$ copies of sets of the form $\mathrm{Div}(z')$ with z' of index at most r. In particular, for $z = \Delta_n^d$, we have:

Corollary 3.6. *For each n and r such that $r < n$, the chain* $(\mathrm{Div}(\Delta_n^d), <)$ *is obtained by concatenating $h_r(\Delta_n^d)$ intervals, each of which, when equipped with $\preccurlyeq$, is a translated copy of some initial sublattice of* $(\mathrm{Div}(\Delta_r^d), \preccurlyeq)$.

The case of Δ_3^2 and Δ_4 are illustrated in Figure 7 and Figure 8.

3B. *Extremal elements.* The next step is to observe that extremal points in B_r-classes admit a simple characterization in terms of divisibility.

Proposition 3.7. *Assume that z is a positive braid.*

(i) *An element x of* $\mathrm{Div}(z)$ *is the maximum of its B_r-class if and only if the relation $x\sigma_i \preccurlyeq z$ fails for $1 \leqslant i < r$.*

(ii) *An element x of* $\mathrm{Div}(z)$ *is the minimum of its B_r-class if and only if no σ_i with $1 \leqslant i < r$ divides x on the right.*

Proof. (i) The condition is necessary: if $x\sigma_i$ lies in $\mathrm{Div}(z)$ for some i with $i < r$, then $x\sigma_i$ lies in the same B_r-class as x, and it is larger both for $\preccurlyeq$ and $<$, so x cannot be maximal in its B_r-class. Conversely, assume that x is not maximal in its B_r-class. Then there exists y satisfying $x < y$ and y is B_r-equivalent to x. Now, by Proposition 3.5, the lcm of x and y is also B_r-equivalent to x, which means that there exists y_1 in B_r^+ satisfying $\mathrm{lcm}(x, y) = xy_1$. Now $x < y$ implies

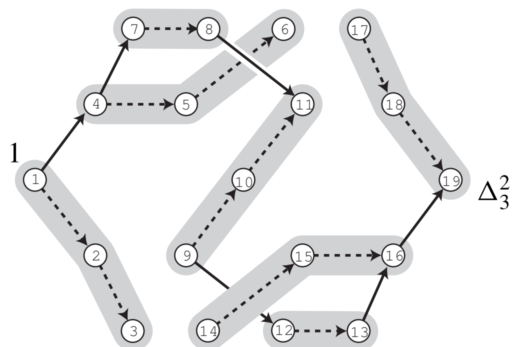


Figure 7. Decomposition of $(\mathrm{Div}(\Delta_3^2), <)$ into B_2-classes. The increasing enumeration of $(\mathrm{Div}(\Delta_3^2), <)$ is the concatenation of the increasing enumeration of the successive classes, separated by 2-jumps (compare with Figure 4); in this case, B_2-classes are simply chains with respect to divisibility.

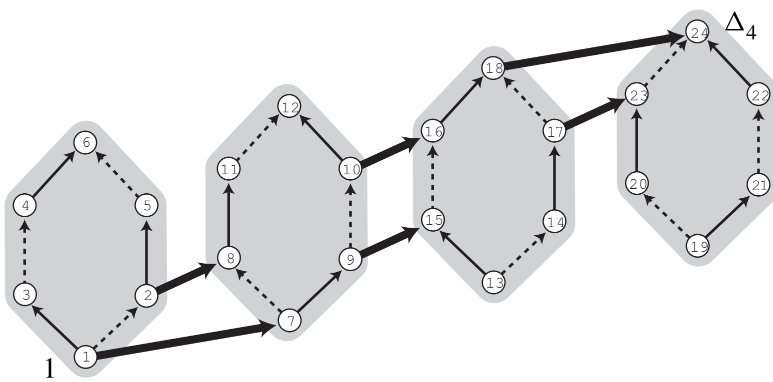


Figure 8. Decomposition of $(\mathrm{Div}(\Delta_4), <)$ into B_3-classes. The σ_3-arrows (thick) corresponding to 3-jumps are not unique; in this case, all B_3-classes are isomorphic to the lattice $(\mathrm{Div}(\Delta_3), <, \preccurlyeq)$, that is, to the Cayley graph of Δ_3.

$y_1 \neq 1$, so there must exist $i < m$ such that σ_i is a left divisor of y_1. Then we have $x\sigma_i \preccurlyeq xy_1 \preccurlyeq z$. Hence $x\sigma_i \preccurlyeq z$.

(ii) The argument is symmetric. If $x = y\sigma_i$ for some positive braid y and $i < r$, then y belongs to the B_r-class of x, and x cannot be minimal in its B_r-class. Conversely, assume that x is not minimal in its B_r-class. Then there exists y satisfying $y < x$ and y is B_r-equivalent to x. By Proposition 3.5 again, the gcd of x and y is also B_r-equivalent to x, which means that there exists y_0 in B_r^+ satisfying $\gcd(x, y)y_0 = x$. As $y < x$ implies $y_0 \neq 1$, there must exist $i < m$ such that σ_i is a right divisor of y_0 and hence of x. □

When we apply the previous criterion to the braids Δ_n^d, we obtain:

Proposition 3.8. *For x in* $\mathrm{Div}(\Delta_n^d)$ *and $1 \leqslant r \leqslant n$, the following are equivalent.*

(i) *The element x is $<$-maximal in its B_r-class.*

(ii) *The element $x\sigma_i$ belongs to* $\mathrm{Div}(\Delta_n^d)$ *for no $i < r$.*

(iii) *The d-th factor of x is right divisible by Δ_r.*

(iv) *The $(d+1)$-st factor of $x\Delta_r$ is Δ_r.*

Proof. The equivalence of (i) and (ii) is given by Proposition 3.7(i). It remains to establish the equivalence of (ii)–(iv). For $r = 1$, (ii) is vacuously true, while (iii) and (iv) always hold. So the expected equivalences are true. We henceforth assume $r \geqslant 2$.

Let x belong to $\mathrm{Div}(\Delta_n^d)$, and let x_d be the d-th factor in the normal form of x. For $i < n$, saying that $x\sigma_i$ does not belong to $\mathrm{Div}(\Delta_n^d)$ means that the normal form of $x\sigma_i$ has length $d+1$. Hence, equivalently, that the normal form of $x_d\sigma_i$ has length 2. This occurs if and only if σ_i is a right divisor of x_d. So, for $r \leqslant n$, (ii) is equivalent to x_d being right divisible by all σ_i's with $1 \leqslant i < r$ and hence to x_d being right divisible by the (left) lcm of these elements, which is Δ_r.

Finally, (iii) and (iv) are equivalent. Indeed, if the d-th factor x_d in the normal form of x is divisible by Δ_r on the right, then (x_d, Δ_r) is a normal sequence as no σ_i with $i < r$ from Δ_r may pass to x_d. Hence $(x_1, \ldots, x_d, \Delta_r)$ is a normal sequence and necessarily the normal form of $x\Delta_r$. Conversely, assume that the normal form of $x\Delta_r$ is $(x_1, \ldots, x_d, \Delta_r)$. The hypothesis that (x_d, Δ_r) is normal implies that x_d is divisible on the right by each σ_i with $i < r$. Hence is divisible on the right by Δ_r. Now $(x_1, \ldots, x_d)$ is the normal form of x. □

Observe that, for $r \geqslant 2$, an element of $\mathrm{Div}(\Delta_n^d)$ that is $<$-maximal in its B_r-class cannot belong to $\mathrm{Div}(\Delta_n^{d-1})$, that is, it cannot have degree $d-1$ or less because the d-th factor of its normal form cannot be 1.

Similar conditions characterize the minimal elements of the B_r-classes. Because the normal form has a privileged orientation, the results are not entirely symmetric with those of Proposition 3.8

Proposition 3.9. *For x in $\mathrm{Div}(\Delta_n^d)$ and $1 \leqslant r \leqslant n$, the following are equivalent.*

(i) *The element x is $<$-minimal in its B_r-class.*

(ii) *No σ_i with $i < r$ is a right divisor of x.*

(iii) *The degrees of x and $x\Delta_r$ are equal.*

Proof. The equivalence of (i) and (ii) is given by Proposition 3.7(ii), and everything is obvious for $r = 1$. So it remains to establish the equivalence of (ii) and (iii) when $r \geqslant 2$. Now, assume that (ii) holds and x has degree d. The hypothesis that σ_i is not a right divisor of x implies that $x\sigma_i$ is a divisor of Δ_n^d. As this holds for each $i < r$, the lcm of $x\sigma_1, \ldots, x\sigma_{r-1}$, namely $x\Delta_r$, also divides Δ_n^d, which means that $x\Delta_r$ has degree (at most) d. So (ii) implies (iii).

Conversely, assume that σ_i divides x on the right. Then the degree of $x\sigma_i$ is strictly larger than that of x, and, *a fortiori*, the same is true for $x\Delta_r$. $\square$

3C. *Determination of $h_r(\Delta_n^d)$.* A direct application of the previous results is a formula connecting the number $h_r(\Delta_n^d)$ of B_r-classes in $\mathrm{Div}(\Delta_n^d)$ with the number of braids whose normal form ends with some specific factor.

Definition 3.10. For $n, d \geqslant 1$ and for s a simple n-braid, we denote by $b_{n,d}(s)$ the number of positive braids of degree at most d, that is, of divisors of Δ_n^d, whose d-th factor is s.

Proposition 3.11. *For $1 \leqslant r \leqslant n$, we have*

$$h_r(\Delta_n^d) = \sum_{s \text{ right divisible by } \Delta_r} b_{n,d}(s) = b_{n,d+1}(\Delta_r). \tag{3-3}$$

In words, the number of r-jumps in $(\mathrm{Div}(\Delta_n^d), <)$ is the number of n-braids of degree at most d whose d-th factor is right divisible by Δ_r.

Proof. By Corollary 3.4, $h_r(\Delta_n^d)$ is the number of B_r-classes in $\mathrm{Div}(\Delta_n^d)$. Each class contains exactly one maximum element, and, by Proposition 3.8, its d-th factor is right divisible by Δ_r. The first equality in (3-3) follows. The second one follows from the equivalence of (iii) and (iv) in Proposition 3.8. $\square$

For $r = 1$, as every simple braid is divisible by 1 on the right, Equation (3-3) reduces to

$$h_1(\Delta_n^d) = \sum_s b_{n,d}(s) = b_{n,d+1}(1),$$

a special case of the relation $h_1(z) = \#\mathrm{Div}(z)$ of Proposition 2.17. For $r = n$, because the only normal sequence of length d that finishes with Δ_n is $(\Delta_n, \dots, \Delta_n)$, Equation (3-3) reduces to

$$h_n(\Delta_n^d) = 1,$$

already noted in Proposition 2.17. Finally, for $r = n-1$, we obtain using Proposition 2.19:

Corollary 3.12. *For $n \geqslant 2$, we have*

$$c(\Delta_n^d) = h_{n-1}(\Delta_n^d) - 1 = \sum_{i=2}^{n} b_{n,d}(\sigma_i\sigma_{i+1}\dots\sigma_{n-1}\Delta_{n-1}) = b_{n,d+1}(\Delta_{n-1}) - 1.$$

Proof. The simple n-braids that are right divisible by Δ_{n-1} are the braids of the form $\sigma_i\sigma_{i+1}\dots\sigma_{n-1}$ with $1 \leqslant i \leqslant n$. Indeed, it is clear that every such braid is simple and right divisible by Δ_{n-1}. Conversely, the only possibility for $z\Delta_{n-1}$ to be simple is that z moves the n-th strand to some position between 1 and n without introducing any crossing between the remaining strands. Finally, $\sigma_1\sigma_2\dots\sigma_{n-1}\Delta_{n-1}$ is Δ_n, and, remembering that $b_{n,d}(\Delta_n)$ is 1, we obtain the first equality. □

3D. *Computation of $b_{n,d}(s)$.* By Lemma 1.4, normal sequences are characterized by a local condition involving only pairs of consecutive elements. It follows that the set of all normal sequences is a rational set, that is, it can be recognized by a finite state automaton. Standard arguments then show that the numbers $b_{n,d}(s)$ obey a linear recurrence. Building on this observation, seemingly first used for braids in [Charney 1995], we can obtain explicit formulas for the parameters $c(\Delta_n^d)$ and $h_r(\Delta_n^d)$ for small values of r, n, or d. We shall not go into details here but refer to [Dehornoy 2007] where we established the formulas and, more generally, investigated the rich combinatorics underlying the normal form of braids.

In the sequel, we write $(M)_{x,y}$ for the (x, y)-entry of a matrix M. The general principle for computing the numbers $b_{n,d}(s)$ for some fixed n is as follows:

Lemma 3.13. *For $n \geqslant 1$, let M_n be the square matrix with entries indexed by simple n-braids defined by*

$$(M_n)_{s,t} = \begin{cases} 1 & \text{if } (s,t) \text{ is normal,} \\ 0 & \text{otherwise.} \end{cases}$$

Then, for every simple t and $d \geqslant 1$, we have $b_{n,d}(t) = ((1, 1, \dots, 1)\, M_n^{d-1})_t$.

The proof is an easy induction on d.

Example 3.14. The matrix M_1 is (1), corresponding to $b_{1,d}(1) = 1$. For $n = 2$, using the enumeration $(1, \sigma_1)$ of simple 2-braids, we find $M_2 = ((1, 0), (1, 1))$, leading to $b_{2,d}(1) = d$ and $b_{2,d}(\sigma_1) = 1$, giving $d + 1$ braids of degree at most

d. The first d are the braids σ_1^e with $e < d$ in which the d-th factor is 1; the last is σ_1^d, whose d-th factor is Δ_2, that is, σ_1. For $n = 3$, using the enumeration $(1, \sigma_1, \sigma_2, \sigma_2\sigma_1, \sigma_1\sigma_2, \Delta_3)$ of simple 3-braids, we obtain

$$M_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

from which we can deduce $b_{3,3}(1) = 19$ or $b_{3,4}(\sigma_1) = 15$ using Lemma 3.13.

Using Proposition 3.11, we deduce:

Proposition 3.15. *With M_n as in Lemma 3.13, we have for $n \geqslant r \geqslant 1$ and $d \geqslant 1$*

$$c(\Delta_n^d) = ((1, 1, \ldots, 1)\, M_n^d)_{\Delta_{n-1}} - 1,$$
$$h_r(\Delta_n^d) = ((1, 1, \ldots, 1)\, M_n^d)_{\Delta_r}.$$

Corollary 3.16. (i) *For fixed n, r, the generating functions for the sequences $c(\Delta_n^d)$ and $h_r(\Delta_n^d)$ are rational.*

(ii) *For fixed n, r, the numbers $c(\Delta_n^d)$ and $h_r(\Delta_n^d)$ admit expressions of the form*

$$P_1(d)\rho_1^d + \cdots + P_k(d)\rho_k^d. \tag{3-4}$$

where $\rho_1, \ldots, \rho_k$ are the nonzero eigenvalues of M_n and $P_1, \ldots, P_k$ are polynomials with $\deg(P_i)$ *of at most the multiplicity of ρ_i in M_n.*

Because the matrix M_n is an $n! \times n!$ matrix, completing the computation is not so easy, even for small values of n. Actually, it is shown in [Dehornoy 2007] how to replace M_n with a smaller matrix $\overline{M}_n$ of size $p(n) \times p(n)$, where $p(n)$ is the number of partitions of n. The property is connected with classical results of Solomon [1976] about the descents of permutations. With such methods, one easily obtains the values listed in Table 1.

Using the reduced matrices

$$\overline{M}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 4 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix} \qquad \text{and} \qquad \overline{M}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 11 & 4 & 1 & 0 & 0 \\ 5 & 3 & 2 & 1 & 0 \\ 6 & 4 & 2 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix},$$

we obtain the following explicit form for (3-4) involving the nonzero eigenvalues $(1, 1, 2)$ of M_3 and $(1, 1, 3 \pm \sqrt{6})$ of M_4:

Proposition 3.17. *Let* $\rho_\pm = 3 \pm \sqrt{6}$. *Then, for* $d \geqslant 1$, *we have*

$$h_1(\Delta_3^d) = 8 \cdot 2^d - 3d - 7,$$
$$h_2(\Delta_3^d) = c(\Delta_3^d) + 1 = 2 \cdot 2^d - 1,$$
$$h_1(\Delta_4^d) = \sum_{\pm} \frac{3}{20}(32 \pm 13\sqrt{6})\rho_\pm^d - \frac{128}{5} \cdot 2^d + 6d + 17,$$
$$h_2(\Delta_4^d) = \sum_{\pm} \frac{1}{20}(32 \pm 13\sqrt{6})\rho_\pm^d - \frac{16}{5} \cdot 2^d + 1,$$
$$h_3(\Delta_4^d) = c(\Delta_3^4) + 1 = \sum_{\pm} \frac{1}{20}(4 \pm \sqrt{6})\rho_\pm^d + \frac{8}{5} \cdot 2^d - 1.$$

These formulas show each parameter grows exponentially in d, with estimate $\mathrm{O}(2^d)$ for $n = 3$, and $\mathrm{O}((3+\sqrt{6})^d)$ for $n = 4$. For practical purposes, it may be more convenient to resort to recursive formulas, for instance,

$$h_1(\Delta_3^d) = 2h_1(\Delta_3^{d-1}) + 3d + 1, \tag{3-5}$$
$$h_1(\Delta_4^d) = 6h_1(\Delta_4^{d-1}) - 3h_1(\Delta_4^{d-2}) + 32 \cdot 2^d - 12d - 34, \tag{3-6}$$

together with initial values $h_1(\Delta_3^0) = h_1(\Delta_4^0) = 1$, $h_1(\Delta_4^1) = 24$ (or $h_1(\Delta_4^{-1}) = 0$).

3E. ***Small values of d.*** Another approach is to keep d fixed and let n vary. Once again, we only mention a few results, and refer the reader to [Dehornoy 2007] for the proofs and additional comments. For $d = 1$, it is easy to determine all values:

Proposition 3.18 [Dehornoy 2007]. *For* $n \geqslant r \geqslant 1$, *we have*

$$h_r(\Delta_n) = \frac{n!}{r!}.$$

For $d = 2$, it is easier to complete the computation for $h_{n-r}(\Delta_n^2)$.

Proposition 3.19 [Dehornoy 2007]. *For* $n \geqslant r \geqslant 1$, *we have*

$$h_{n-r}(\Delta_n^2) = r!\,(r+1)^n + \sum_{i=1}^{r} P_i(n)\, i^{n-r+i-1},$$

for some polynomial P_i *of degree at most* $r - i + 1$. *The values for* $r = 1, 2$ *are*

$$h_{n-1}(\Delta_n^2) = 2^n - 1,$$
$$h_{n-2}(\Delta_n^2) = 2 \cdot 3^n - (n+6) \cdot 2^{n-1} + 1.$$

For $h_r(\Delta_n^2)$ itself, no general formula is known. We mention the case of $h_1(\Delta_n^2)$, which follows from results of Carlitz et al. [1976]:

Proposition 3.20 [Dehornoy 2007]. *The numbers $h_1(\Delta_n^2)$ are determined by the induction*

$$h_1(\Delta_0^2) = 1, \qquad h_1(\Delta_n^2) = \sum_{i=0}^{n-1} (-1)^{n+i+1} \binom{n}{i}^2 h_1(\Delta_i^2).$$

Their double exponential generating function is, *with $J_0(x)$ is the Bessel function*,

$$\sum_{n=0}^{\infty} h_1(\Delta_n^2) \frac{z^n}{n!^2} = \left(\sum_{n=0}^{\infty} (-1)^n \frac{z^n}{n!^2} \right)^{-1} = \frac{1}{J_0(\sqrt{z})}.$$

Finally, for $d = 3$, the computation can be completed at least in the case $n - r = 1$:

Proposition 3.21 [Dehornoy 2007]. *For $n \geqslant 1$*, *we have*, *with $e = \exp(1)$*,

$$h_{n-1}(\Delta_n^3) = \sum_{i=0}^{n-1} \frac{n!}{i!} = \lfloor n!e \rfloor - 1.$$

d	0	1	2	3	4	5	6
$h_1(\Delta_2^d)$	1	2	3	4	5	6	7
$h_1(\Delta_3^d)$	1	6	19	48	109	234	487
$h_2(\Delta_3^d)$	1	3	7	15	31	63	127
$h_1(\Delta_4^d)$	1	24	211	1,380	8,077	45,252	249,223
$h_2(\Delta_4^d)$	1	12	83	492	2,765	15,240	83,399
$h_3(\Delta_4^d)$	1	4	15	64	309	1,600	8,547
$h_1(\Delta_5^d)$	1	120	3,651	79,140	1,548,701	29,375,460	551,997,751
$h_2(\Delta_5^d)$	1	60	1,501	30,540	585,811	11,044,080	207,154,921
$h_3(\Delta_5^d)$	1	20	311	5,260	94,881	1,755,360	32,741,851
$h_4(\Delta_5^d)$	1	5	31	325	4,931	86,565	1,590,231
$h_1(\Delta_6^d)$	1	720	90,921	7,952,040	634,472,921	49,477,263,360	3,836,712,177,121
$h_2(\Delta_6^d)$	1	360	38,559	3,228,300	254,718,389	19,808,530,620	1,535,016,069,499
$h_3(\Delta_6^d)$	1	120	8,727	649,260	49,654,757	3,831,626,580	296,361,570,667
$h_4(\Delta_6^d)$	1	30	1,075	61,620	4,387,195	332,578,230	25,612,893,355
$h_5(\Delta_6^d)$	1	6	63	1,956	116,423	8,448,606	643,888,543

Table 1. First values of $h_r(\Delta_n^d)$ for $1 \leqslant r < n$ — the value is 1 for $r \geqslant n$. For instance, the number $h_1(\Delta_3^2)$ of 3-strand braids of degree at most 2 is 19 (see Example 2.16), while the maximal number $c(\Delta_4^4)$ of σ_3's in a σ-positive word drawn in $\Gamma(\Delta_4^4)$ — which is $h_3(\Delta_4^4) - 1$, according to Proposition 2.19 — is 308.

Using Proposition 2.19, we deduce the following explicit values for $c(\Delta_n^d)$, that is, for the maximal number of occurrences of σ_{n-1} in a σ-positive word drawn in the Cayley graph of Δ_n^d:

$$c(\Delta_n) = n-1, \qquad c(\Delta_n^2) = 2^n - 2, \qquad c(\Delta_n^3) = \sum_{i=0}^{n-1} \frac{n!}{i!} - 1 = \lfloor n!e \rfloor - 2.$$

The formulas listed above show that a number of different induction schemes appear, suggesting that the combinatorics of normal sequences of braids is very rich.

4. A complete description of $(\mathrm{Div}(\Delta_3^d), <)$

Our ultimate goal is a complete description of each chain $(\mathrm{Div}(\Delta_n^d), <)$. Typically, this means that we are able to explicitly specify the increasing enumeration of its elements. The goal remains generally out of reach, but we can show how the process can be completed when $n = 3$. The counting formulas of Section 3 play a key role in the construction, and, in particular, the Pascal's triangle of Figure 9 connects directly with the 2^d factor in the inductive formulas of Proposition 3.17. As an application, we deduce a new proof of Property C and of the well-ordering property and hence a complete reconstruction of the braid ordering when $n = 3$.

The general principle is to make the decomposition of Corollary 3.6 explicit. The latter shows that, for all n and d, the chain $(\mathrm{Div}(\Delta_n^d), <)$ can be decomposed into $c(\Delta_n^d)$ subintervals each of which copies some fragment of $(\mathrm{Div}(\Delta_{n-1}^d), <)$. Moreover, the approach of Section 3 suggests an induction on d as well. We are led to seek a recursion for $(\mathrm{Div}(\Delta_n^d), <)$ in $(\mathrm{Div}(\Delta_{n-1}^d), <)$ and $(\mathrm{Div}(\Delta_n^{d-1}), <)$; here this means expressing $(\mathrm{Div}(\Delta_3^d), <)$ in $(\mathrm{Div}(\Delta_2^d), <)$ and $(\mathrm{Div}(\Delta_3^{d-1}), <)$.

4A. ***The braids $\theta_{n,p}$.*** The subsequent construction will appeal to a double series $\theta_{n,p}$ of braids, and we begin with a few preliminary properties.

Definition 4.1. For $n \geqslant 2$, let $\sigma_{n,1}$ and $\sigma_{1,n}$ denote the braid words $\sigma_{n-1}\sigma_{n-2}\dots\sigma_1$ and $\sigma_1\sigma_2\dots\sigma_{n-1}$. For $p \geqslant 0$, we define $\widetilde{\theta}_{n,p}$ as (the braid represented by) the length p prefix of the right-infinite word $(\sigma_{n,1}\sigma_{1,n})^\infty$, and let $\theta_{n,p}$ be (the braid represented by) the length p suffix of the left-infinite word ${}^\infty(\sigma_{n,1}\sigma_{1,n})$.

For instance, we find $\theta_{3,0} = 1$, $\theta_{3,1} = \mathtt{b}$, $\theta_{3,2} = \mathtt{ab}$, $\dots$, $\theta_{3,4} = \mathtt{baab}$, $\dots$, $\theta_{3,7} = \mathtt{aabbaab}$, etc. Similarly, we have $\theta_{4,6} = \mathtt{cbaabc}$ and, more generally, $\theta_{n,2n-2} = \widetilde{\theta}_{n,2n-2} = \sigma_{n,1}\sigma_{1,n}$. Note that, as words, $\theta_{n,p}$ is the reverse of $\widetilde{\theta}_{n,p}$.

Lemma 4.2. *For $n \geqslant 2$ and $p, q \geqslant 0$ satisfying $p + q = d(n-1)$, we have*

$$\theta_{n,p}\, \Delta_{n-1}^d\, \widetilde{\theta}_{n,q} = \Delta_n^d. \tag{4-1}$$

Proof. We first prove using induction on d the relation

$$\theta_{n,d(n-1)}\,\Delta_{n-1}^{d} = \Delta_{n}^{d}, \tag{4-2}$$

that is, (4-1) with $q = 0$. For $d = 0$, (4-2) reduces to $1 = 1$. Assume $d \geqslant 1$. By definition, $\theta_{n,d(n-1)}$ is $\sigma_{n,1}\,\theta_{n,(d-1)(n-1)}$ for d odd and is $\sigma_{1,n}\,\theta_{n,(d-1)(n-1)}$ for d even. In either case, we can write

$$\theta_{n,d(n-1)} = \phi_n^{d-1}(\sigma_{1,n})\,\theta_{n,(d-1)(n-1)},$$

where we recall ϕ_n denotes the flip automorphism of B_n that exchanges σ_i and σ_{n-i}. Using the induction hypothesis and (1-2), we find

$$\begin{aligned}\theta_{n,d(n-1)}\,\Delta_{n-1}^{d} &= \phi_n^{d-1}(\sigma_{1,n})\,\theta_{n,(d-1)(n-1)}\,\Delta_{n-1}^{d-1}\,\Delta_{n-1}\\ &= \phi_n^{d-1}(\sigma_{1,n})\,\Delta_{n}^{d-1}\,\Delta_{n-1} = \Delta_{n}^{d-1}\,\sigma_{1,n}\,\Delta_{n-1} = \Delta_{n}^{d-1}\,\Delta_{n} = \Delta_{n}^{d}.\end{aligned}$$

We return to the general case of (4-1). For d even, we have $\theta_{n,d(n-1)} = \widetilde{\theta}_{n,d(n-1)}$ and hence $\widetilde{\theta}_{n,q}\,\theta_{n,p} = \theta_{n,d(n-1)}$. If d is odd, we have $\theta_{n,d(n-1)} = \phi_n(\widetilde{\theta}_{n,d(n-1)})$, which implies $\phi_n(\widetilde{\theta}_{n,q})\,\theta_{n,p} = \theta_{n,d(n-1)}$. So $\phi_n^{d}(\widetilde{\theta}_{n,q})\,\theta_{n,p} = \theta_{n,d(n-1)}$ holds in both cases. Now, using (4-2), we find

$$\phi_n(\widetilde{\theta}_{n,q})\,\theta_{n,p}\,\Delta_{n-1}^{d}\,\widetilde{\theta}_{n,q} = \theta_{n,d(n-1)}\,\Delta_{n-1}^{d}\,\widetilde{\theta}_{n,q} = \Delta_{n}^{d}\,\widetilde{\theta}_{n,q} = \phi_n(\widetilde{\theta}_{n,q})\,\Delta_{n}^{d},$$

from which we deduce (4-1) by cancelling $\phi_n(\widetilde{\theta}_{n,q})$ on the left. □

Lemma 4.3. *For* $1 \leqslant i \leqslant n-2$ *we have*

$$\theta_{n,d(n-1)}\,\sigma_i = \sigma_{i+e}\,\theta_{n,d(n-1)} \tag{4-3}$$

with $e = 0$ *if* d *is even and* $e = 1$ *if* d *is odd.*

Proof. For $1 \leqslant i \leqslant n-2$, we have $\sigma_{1,n}\,\sigma_i = \sigma_{i+1}\,\sigma_{1,n}$ and $\sigma_{n,1}\,\sigma_{i+1} = \sigma_i\,\sigma_{n,1}$, as an easy induction shows. This implies $\sigma_{n,1}\,\sigma_{1,n}\,\sigma_i = \sigma_i\,\sigma_{n,1}\,\sigma_{1,n}$ and therefore $(\sigma_{n,1}\,\sigma_{1,n})^d\,\sigma_i = \sigma_i\,(\sigma_{n,1}\,\sigma_{1,n})^d$, that is, $\theta_{n,2d(n-1)}\,\sigma_i = \sigma_i\,\theta_{n,2d(n-1)}$ for every d. On the other hand, we have $\theta_{n,(2d+1)(n-1)} = \sigma_{1,n}\,\theta_{n,2d(n-1)}$ and hence

$$\theta_{n,(2d+1)(n-1)}\,\sigma_i = \sigma_{1,n}\,\sigma_i\,\theta_{n,2d(n-1)} = \sigma_{i+1}\,\sigma_{1,n}\,\theta_{n,2d(n-1)} = \sigma_{i+1}\,\theta_{n,(2d+1)(n-1)},$$

as was expected. □

4B. ***A Pascal triangle.*** We shall now construct for every d a sequence of positive braids S_3^d that will be the increasing enumeration of $(\mathrm{Div}(\Delta_3^d), <)$. The construction relies on an induction similar to Pascal's triangle. To make it easily understandable, we start with the (trivial) cases $n = 1$ and $n = 2$.

Because B_1 is the trivial group, for every d, 1 is the only element of degree at most d, and we can state:

Proposition 4.4. *Define S_1^d for $d \geqslant 0$ by*

$$S_1^d = (1).$$

Then S_1^d is the increasing enumeration of $\mathrm{Div}(\Delta_1^d)$.

The group B_2 is the rank 1 free group generated by σ_1. The braid Δ_2 is just σ_1, and the braids of degree at most d, that is, the divisors of Δ_2^d, consist of the $d+1$ braids $1, \sigma_1, \dots, \sigma_1^d$. On the other hand, we have $\sigma_{1,2} = \sigma_{2,1} = \sigma_1$, and $\theta_{1,i} = \sigma_1^i$ for every i.

Notation 4.5. If S_1, S_2 are sequences (of braids), we denote by $S_1 + S_2$ the (ordered) concatenation of S_1 and S_2. If S is a sequence of braids and x is a braid, we denote by xS the translated sequence obtained by multiplying each entry in S by x on the left.

In these terms, the sequence $(1, \sigma_1, \dots, \sigma_1^d)$ can be expressed as a sum of sequences $\theta_{2,0}(1) + \theta_{2,1}(1) + \cdots + \theta_{2,d}(1)$. Hence:

Proposition 4.6. *Define S_2^d for $d \geqslant 0$ by*

$$S_2^d = \theta_{2,0} S_1^d + \theta_{2,1} S_1^d + \cdots + \theta_{2,d} S_1^d. \tag{4-4}$$

Then S_2^d is the increasing enumeration of $\mathrm{Div}(\Delta_2^d)$.

We repeat the process for $n = 3$, introducing a sequence S_3^d by a definition similar to (4-4) that involves S_2^d and S_3^{d-1}. The result we shall prove is:

Proposition 4.7. *Let S_3^d be defined for $d \geqslant 0$ by*

$$S_3^d = \theta_{3,0} S_2^d + S_3^{d,1} + \theta_{3,1} S_2^d + \cdots + \theta_{3,2d-1} S_2^d + S_3^{d,2d} + \theta_{3,2d} S_2^d, \tag{4-5}$$

where $S_3^{d,1}, \cdots, S_3^{d,2d}$ are defined by $S_3^{d,1} = S_3^{d,2d} = \varnothing$ and, for $2 \leqslant p \leqslant 2d-1$,

$$S_3^{d,p} = \begin{cases} \sigma_1 (S_3^{d-1,p-1} + \theta_{3,p-1} S_2^{d-1} + S_3^{d-1,p}) & \text{for } p = 0 \pmod 4, \\ \sigma_2\sigma_1 (S_3^{d-1,p-2} + \theta_{3,p-1} S_2^{d-1} + S_3^{d-1,p-1}) & \text{for } p = 1 \pmod 4, \\ \sigma_2 (S_3^{d-1,p-1} + \theta_{3,p-1} S_2^{d-1} + S_3^{d-1,p}) & \text{for } p = 2 \pmod 4, \\ \sigma_1\sigma_2 (S_3^{d-1,p-2} + \theta_{3,p-1} S_2^{d-1} + S_3^{d-1,p-1}) & \text{for } p = 3 \pmod 4. \end{cases}$$

Then S_3^d is the increasing enumeration of $\mathrm{Div}(\Delta_3^d)$.

The general scheme is illustrated in Figure 9. The sequence S_3^d is constructed by starting with $2d+1$ copies of S_2^d translated by $\theta_{3,0}, \dots, \theta_{3,2d}$ and inserting (translated copies of) fragments of the previous sequence S_3^{d-1}.

Example 4.8. The difference between the definition of S_3^d in (4-5) and that of S_2^d in (4-4) is the insertion of the additional factors $S_3^{d,p}$ between the consecutive terms

$$\theta_{3,0}S_2^0$$

$$\theta_{3,0}S_2^1\ \underbrace{(S_3^{1,1})\ \theta_{3,1}S_2^1\ (S_3^{1,2})}\ \theta_{3,2}S_2^1$$

$$\sigma_2\cdot\swarrow \qquad \searrow\sigma_1\sigma_2\cdot$$

$$\theta_{3,0}S_2^2\ \underbrace{(S_3^{2,1})\ \theta_{3,1}S_2^2\ \ S_3^{2,2}}\ \theta_{3,2}S_2^2\ \ \underbrace{S_3^{2,3}\ \ \theta_{3,3}S_2^2\ (S_3^{2,4})}\ \theta_{3,4}S_2^2$$

$$\sigma_2\cdot\swarrow \qquad \searrow\sigma_1\sigma_2\cdot \qquad\qquad \sigma_1\cdot\swarrow \qquad \searrow\sigma_2\sigma_1\cdot$$

$$\theta_{3,0}S_2^3\ \underbrace{(S_3^{3,1})\ \theta_{3,1}S_2^3\ \ S_3^{3,2}}\ \theta_{3,2}S_2^3\ \ \underbrace{S_3^{3,3}\ \ \theta_{3,3}S_2^3\ \ S_3^{3,4}}\ \theta_{3,4}S_2^3\ \ \underbrace{S_3^{3,5}\ \ \theta_{3,4}S_2^3\ (S_3^{3,6})}\ \theta_{3,6}S_2^3$$

$$\sigma_2\cdot\swarrow \quad \searrow\sigma_1\sigma_2\cdot \qquad \sigma_1\cdot\swarrow \quad \searrow\sigma_2\sigma_1\cdot \qquad \sigma_2\cdot\swarrow \quad \searrow\sigma_1\sigma_2\cdot$$

$$\dots \quad \dots \qquad \dots \quad \dots \qquad \dots \quad \dots$$

Figure 9. The inductive construction of S_3^d as a Pascal triangle: the subsequence $S_3^{d,p}$ is obtained by translating and concatenating the previous subsequences $S_3^{d-1,p-1}$ and $S_3^{d-1,p}$, or $S_3^{d-1,p-2}$ and $S_3^{d-1,p-1}$, depending on the parity of p. The bracketed sequences are empty; if we remove the subsequences $\theta_{3,q}S_2^d$, we have the Pascal triangle.

$\theta_{3,q}S_2^d$. Because $S_3^{d,1}$ and $S_3^{d,2d}$ are empty, the difference occurs for $d \geqslant 2$ only. The first values are:

$$\begin{aligned}
S_3^0 &= \theta_{3,0}S_2^0 = (1),\\
S_3^1 &= \theta_{3,0}S_2^1 + S_3^{1,1} + \theta_{3,1}S_2^1 + S_3^{1,2} + \theta_{3,2}\\
&= (1, \mathtt{a}) + \varnothing + \mathtt{b}(1, \mathtt{a}) + \varnothing + \mathtt{ab}(1, \mathtt{a}) = (1, \mathtt{a}, \mathtt{b}, \mathtt{ba}, \mathtt{ab}, \mathtt{aba}),\\
S_3^2 &= \theta_{3,0}S_2^2 + S_3^{2,1} + \theta_{3,1}S_2^2 + S_3^{2,2} + \theta_{3,2}S_2^2 + S_3^{2,3} + \theta_{3,3}S_2^2 + S_3^{2,4} + \theta_{3,4}S_2^2\\
&= (1, \mathtt{a}, \mathtt{aa}) + \varnothing + \mathtt{b}(1, \mathtt{a}, \mathtt{aa}) + \mathtt{b}(\mathtt{b}, \mathtt{ba}) + \mathtt{ab}(1, \mathtt{a}, \mathtt{aa})\\
&\qquad\qquad + \mathtt{ab}(\mathtt{b}, \mathtt{ba}) + \mathtt{aab}(1, \mathtt{a}, \mathtt{aa}) + \varnothing + \mathtt{baab}(1, \mathtt{a}, \mathtt{aa})\\
&= (1, \mathtt{a}, \mathtt{aa}, \mathtt{b}, \mathtt{ba}, \mathtt{baa}, \mathtt{bb}, \mathtt{bba}, \mathtt{ab}, \mathtt{aba}, \mathtt{abaa}, \mathtt{abb}, \mathtt{abba}, \mathtt{aab},\\
&\qquad\qquad \mathtt{aaba}, \mathtt{aabaa}, \mathtt{baab}, \mathtt{baaba}, \mathtt{baabaa}).
\end{aligned}$$

It is easy to check directly that the sequence S_3^d provides the increasing enumeration of $\mathrm{Div}(\Delta_3^d)$ for $d = 0, 1, 2$.

The proof of Proposition 4.7 will be split into several pieces, each of which is established using an induction on the degree d.

Lemma 4.9. *All entries in S_3^d are divisors of Δ_3^d.*

Proof. The result is true for $d = 0$. Assume $d \geqslant 1$. By construction, each entry in S_3^d either is of the form $\theta_{3,q}\sigma_1^e$ with $0 \leqslant q \leqslant 2d$ and $0 \leqslant e \leqslant d$ or belongs to some subsequence $S_3^{d,p}$ with $2 \leqslant p \leqslant 2d-1$. In the first case, $\theta_{3,q}\sigma_1^e$ is a right divisor of $\theta_{3,2d}\sigma_1^e$, which itself is a left divisor of $\theta_{3,2d}\sigma_1^d$. By Equation (4-1), the latter is Δ_3^d. Hence each $\theta_{3,q}\sigma_1^e$ is a divisor of Δ_3^d. As for the entries coming from some subsequence $S_3^{d,p}$, by definition they are of the form xy with x one of $\sigma_2, \sigma_1\sigma_2, \sigma_1, \sigma_2\sigma_1$ and y an entry in S_3^{d-1}. Then x is a divisor of Δ_3, while, by the induction hypothesis, y is a divisor of Δ_3^{d-1}. Thus xy is a divisor of Δ_3^d. □

Lemma 4.10. *The length of the sequence S_3^d equals the cardinality of* $\mathrm{Div}(\Delta_3^d)$.

Proof. Let ℓ_d denote the length of S_3^d. Computing ℓ_d by recursion is not very difficult but also unnecessary. Indeed, we saw in Section 3 that the cardinality $h_1(\Delta_3^d)$ of $\mathrm{Div}(\Delta_3^d)$ obeys the inductive rule (3-5). So it will be enough to check that ℓ_d satisfies the relation

$$\ell_d = 2\ell_{d-1} + 3d + 1 \tag{4-6}$$

and starts from the initial $\ell_1 = 6$ (or $\ell_0 = 1$). The latter point was checked in Example 4.8.

Figure 9 shows that most entries in S_3^{d-1} generate two entries in S_3^d. More precisely, each entry of S_3^{d-1} not belonging to a factor of the form $\theta_{3,2q}S_2^{d-1}$ generates two entries in S_3^d, and, conversely, each entry in S_3^d not belonging to a factor $\theta_{3,q}S_2^d$ comes from such an entry in S_3^{d-1}. The d factors $\theta_{3,2q}S_2^{d-1}$ in S_3^{d-1} each have length d, and the $2d+1$ factors $\theta_{3,2q}S_2^d$ in S_3^d each have length $d+1$. So we obtain

$$\ell_d - (2d+1)(d+1) = 2(\ell_{d-1} - d^2),$$

which gives Equation (4-6). □

At this point, we cannot (yet) conclude that each divisor of Δ_3^d occurs exactly once in S_3^d, as there could be some repetitions.

4C. *A quotient sequence for S_3^d.* Our next aim is to show that S_3^d is $<$-increasing. To this end, we shall explicitly determine the quotient of adjacent entries in S_3^d, that is, we shall specify a quotient sequence for S_3^d in the sense of Definition 2.22.

We begin by determining the first and the last entries of the sequence $S_3^{d,p}$. For S a nonempty sequence, we denote by $(S)_1$ and $(S)_\infty$ the first and last entry in S.

Lemma 4.11. *For $1 < p < 2d$, we have*

$$(S_3^{d,p})_1 = \theta_{3,p-1}\,\sigma_2 \quad \textit{and} \quad (S_3^{d,p})_\infty\,\sigma_2 = \theta_{3,p}\,\sigma_1^d.$$

Proof. The result is vacuously true for $d=0, 1$. Assume $d \geqslant 2$ with $p=0 \pmod 4$. Using the definition, the induction hypothesis, and (4-3), we find

$$\begin{aligned}(S_3^{d,p})_1 &= \sigma_1\,(S_3^{d-1,p-1})_1 = \sigma_1\,\theta_{3,p-2}\,\sigma_2 = \theta_{3,p-1}\,\sigma_2,\\ (S_3^{d,p})_\infty\,\sigma_2 &= \sigma_1\,(S_3^{d-1,p})_\infty\,\sigma_2 = \sigma_1\,\theta_{3,p}\,\sigma_1^{d-1} = \theta_{3,p}\,\sigma_1^d.\end{aligned}$$

Similarly, for $p = 1 \pmod 4$, we have

$$(S_3^{d,p})_1 = \sigma_2\sigma_1\,(S_3^{d-1,p-2})_1 = \sigma_2\sigma_1\,\theta_{3,p-3}\,\sigma_2 = \theta_{3,p-1}\,\sigma_2,$$
$$(S_3^{d,p})_\infty\,\sigma_2 = \sigma_2\sigma_1\,(S_3^{d-1,p-1})_\infty\,\sigma_2 = \sigma_2\sigma_1\,\theta_{3,p-1}\,\sigma_1^{d-1} = \sigma_2\,\theta_{3,p-1}\,\sigma_1^d = \theta_{3,p}\,\sigma_1^d.$$

Then, for $p = 2 \pmod 4$, we have

$$(S_3^{d,p})_1 = \sigma_2\,(S_3^{d-1,p-1})_1 = \sigma_2\,\theta_{3,p-2}\,\sigma_2 = \theta_{3,p-1}\,\sigma_2,$$
$$(S_3^{d,p})_\infty\,\sigma_2 = \sigma_2\,(S_3^{d-1,p})_\infty\,\sigma_2 = \sigma_2\,\theta_{3,p}\,\sigma_1^{d-1} = \theta_{3,p}\,\sigma_1^d.$$

Finally, for $p = 3 \pmod 4$, we find

$$(S_3^{d,p})_1 = \sigma_1\sigma_2\,(S_3^{d-1,p-2})_1 = \sigma_1\sigma_2\,\theta_{3,p-3}\,\sigma_2 = \theta_{3,p-1}\,\sigma_2,$$
$$\begin{aligned}(S_3^{d,p})_\infty\,\sigma_2 &= \sigma_1\sigma_2\,(S_3^{d-1,p-1})_\infty\,\sigma_2 = \sigma_1\sigma_2\,\theta_{3,p-1}\,\sigma_1^{d-1} = \sigma_1\sigma_2\sigma_1\sigma_2\,\theta_{3,p-3}\,\sigma_1^{d-1}\\ &= \sigma_1\sigma_1\sigma_2\sigma_1\,\theta_{3,p-3}\,\sigma_1^{d-1} = \sigma_1\sigma_1\sigma_2\,\theta_{3,p-3}\,\sigma_1^d = \theta_{3,p}\,\sigma_1^d.\end{aligned}$$ □

We shall now construct an explicit quotient sequence for S_3^d, that is, a sequence of braid words representing the quotients of the consecutive entries of S_3^d. Before doing it for S_3^d, let us consider the (trivial) cases of S_1^d and S_2^d. As S_1^d consists of one single entry, it vacuously admits the empty sequence as a quotient sequence. As for S_2^d, we can state:

Lemma 4.12. *For $d \geqslant 0$, let $\boldsymbol{w}_1^d$ be the empty sequence, and let $\boldsymbol{w}_2^d$ be defined by*

$$\boldsymbol{w}_2^d = \boldsymbol{w}_1^d + (\sigma_1) + \boldsymbol{w}_1^d + \cdots + \boldsymbol{w}_1^d + (\sigma_1) + \boldsymbol{w}_1^d,$$

d times (σ_1). Then $\boldsymbol{w}_2^d$ is a quotient sequence for S_2^d.

In a similar way, we shall prove:

Proposition 4.13. *Let $\boldsymbol{w}_3^d$ be the sequence defined by $\boldsymbol{w}_3^0 = \varnothing$ and*

$$\begin{aligned}\boldsymbol{w}_3^d = \boldsymbol{w}_2^d + (\sigma_1^{-d}\sigma_2) &+ \boldsymbol{w}_2^d + (\sigma_1^{-d}\sigma_2) + \boldsymbol{w}_3^{d,2} + (\sigma_2\sigma_1^{-d})\\ &+ \boldsymbol{w}_2^d + (\sigma_1^{-d}\sigma_2) + \boldsymbol{w}_3^{d,3} + (\sigma_2\sigma_1^{-d}) + \cdots\\ &+ \boldsymbol{w}_2^d + (\sigma_1^{-d}\sigma_2) + \boldsymbol{w}_3^{d,2d-1} + (\sigma_2\sigma_1^{-d})\\ &\qquad + \boldsymbol{w}_2^d + (\sigma_2\sigma_1^{-d}) + \boldsymbol{w}_2^d,\end{aligned} \tag{4-7}$$

with

$$\boldsymbol{w}_3^{d,2} = \boldsymbol{w}_3^{d,3} = \boldsymbol{w}_2^{d-1} + (\sigma_2\sigma_1^{-d+1}) + \boldsymbol{w}_3^{d-1,2},$$
$$\boldsymbol{w}_3^{d,2d-2} = \boldsymbol{w}_3^{d,2d-1} = \boldsymbol{w}_3^{d-1,2d-3} + (\sigma_1^{-d+1}\sigma_2) + \boldsymbol{w}_2^{d-1},$$
$$\boldsymbol{w}_3^{d,2p} = \boldsymbol{w}_3^{d,2p+1} = \boldsymbol{w}_3^{d-1,2p-1} + (\sigma_1^{-d+1}\sigma_2) + \boldsymbol{w}_2^{d-1} + (\sigma_2\sigma_1^{-d+1}) + \boldsymbol{w}_3^{d-1,2p},$$

for $4 \leqslant 2p \leqslant 2d-4$. Then $\boldsymbol{w}_3^d$ is a quotient sequence for S_3^d.

Example 4.14. We find $\boldsymbol{w}_3^1 = \boldsymbol{w}_2^1 + (\mathtt{Ab}) + \boldsymbol{w}_2^1 + (\mathtt{bA}) + \boldsymbol{w}_2^1 = (\mathtt{a}, \mathtt{Ab}, \mathtt{a}, \mathtt{bA}, \mathtt{a})$, and

$$\begin{aligned}\boldsymbol{w}_3^2 = \boldsymbol{w}_2^2 &+ (\mathtt{AAb}) + \boldsymbol{w}_2^2 + (\mathtt{AAb}) + \boldsymbol{w}_3^{2,2} + (\mathtt{bAA}) \\ &+ \boldsymbol{w}_2^2 + (\mathtt{AAb}) + \boldsymbol{w}_3^{2,3} + (\mathtt{bAA}) + \boldsymbol{w}_2^2 + (\mathtt{bAA}) + \boldsymbol{w}_2^2\end{aligned}$$

with $\boldsymbol{w}_3^{2,2} = \boldsymbol{w}_3^{2,3} = \boldsymbol{w}_2^1 = (\mathtt{a})$, whence

$$\boldsymbol{w}_3^2 = (\mathtt{a}, \mathtt{a}, \mathtt{AAb}, \mathtt{a}, \mathtt{a}, \mathtt{AAb}, \mathtt{a}, \mathtt{bAA}, \mathtt{a}, \mathtt{a}, \mathtt{AAb}, \mathtt{a}, \mathtt{bAA}, \mathtt{a}, \mathtt{a}, \mathtt{bAA}, \mathtt{a}, \mathtt{a}).$$

Proof of Proposition 4.13. We prove using induction on d that $\boldsymbol{w}_3^d$ is a quotient sequence for S_3^d with the $4d-2$ terms in (4-7) corresponding to the $4d-1$ nonempty terms in (4-5). In particular, for $2 \leqslant p \leqslant 2d-1$, the subsequence $\boldsymbol{w}_3^{d,p}$ is a quotient sequence for $S_3^{d,p}$. The result is vacuously true for $d = 0$. Assume $d \geqslant 1$. By definition, the sequence S_3^d consists of the concatenation of the $2d+1$ sequences $\theta_{3,0}S_2^d, \cdots, \theta_{3,2d}S_2^d$, in which the $2d-2$ sequences $S_3^{d,2}, \ldots, S_3^{d,2d-1}$ are inserted. We shall consider these subsequences separately and then consider the transitions between consecutive subsequences.

First, since $\boldsymbol{w}_2^d$ is a quotient sequence for S_2^d, it is a quotient sequence for every sequence $\theta_{3,q}S_2^d$ as well, because, by definition, the quotients we consider are invariant under left translation. Then, by construction, each subsequence $S_3^{d,2p}$ or $S_3^{d,2p+1}$ appearing in S_3^d is obtained by translating some subsequence S of S_3^{d-1}, namely

$$S = S_3^{d-1,2p-1} + \theta_{3,q-1}S_2^{d-1} + S_3^{d-1,2p}.$$

By the induction hypothesis, the sequence

$$\boldsymbol{w}_3^{d-1,2p-1} + (\sigma_1^{-d+1}\sigma_2) + \boldsymbol{w}_2^{d-1} + (\sigma_2\sigma_1^{-d+1}) + \boldsymbol{w}_3^{d-1,2p},$$

which by definition is precisely $\boldsymbol{w}_3^{d,2p}$ and $\boldsymbol{w}_3^{d,2p+1}$, is a quotient sequence for S. The property remains true in the special cases $p = 1$ and $p = d$, which correspond respectively to removing the initial term $S_3^{d-1,2p-1}$ and the final term $S_3^{d-1,2p}$. Then $\boldsymbol{w}_3^{d,2p}$ and $\boldsymbol{w}_3^{d,2p+1}$ are also quotient sequences for any sequence obtained from S by a left translation, and, in particular, for $S_3^{d,2p}$ and $S_3^{d,2p+1}$.

It remains to study the transitions between the consecutive terms in the expression (4-5) of S_3^d, that is, to compare the last entry in each term with the first entry in the next term. Four cases are to be considered, namely the special cases of the first two terms and of the final two terms, and the generic cases of the transitions from $\theta_{3,q}S_2^d$ to $S_3^{d,p+1}$ and from $S_3^{d,p}$ to $\theta_{3,q}S_2^d$.

As for the first two terms $\theta_{3,0}S_2^d = S_2^d$ and $\theta_{3,1}S_2^d = \sigma_2 S_2^d$, the last entry in S_2^d is σ_1^d, while the first entry in $\sigma_2 S_2^d$ is σ_2, so $\sigma_1^{-d}\sigma_2$ is a quotient. For the last two terms $\theta_{3,2d-1}S_2^d$ and $\theta_{3,2d}S_2^d$, the last entry in $\theta_{3,2d-1}S_2^d$ is $\theta_{3,2d-1}\,\sigma_1^d$, while the first entry in $\theta_{3,2d}S_2^d$ is $\theta_{3,2d}$. Now, by (4-1), we have $\theta_{3,2d-1}\,\sigma_1^d\sigma_2 = \theta_{3,2d}\,\sigma_1^d$, so $\sigma_2\sigma_1^{-d}$ expresses the quotient.

Consider now the transition from $\theta_{3,q} S_2^d$ to $S_3^{d,q+1}$. The last entry in $\theta_{3,q} S_2^d$ is $\theta_{3,q}\,\sigma_1^d$, while, by Lemma 4.11, the first entry in $S_3^{d,q+1}$ is $\theta_{3,q}\,\sigma_2$. Hence $\sigma_1^{-d}\sigma_2$ represents the quotient. Finally, consider the transition from $S_3^{d,p}$ to $\theta_{3,q} S_2^d$. By Lemma 4.11 again, the last entry x in $\theta_{3,q} S_2^d$ satisfies $x\,\sigma_2 = \theta_{3,q}\,\sigma_1^d$, while the first entry in $\theta_{3,q} S_2^d$ is $\theta_{3,q}$. Hence $\sigma_2\sigma_1^{-d}$ represents the quotient. □

Corollary 4.15. *For each d the sequence S_3^d is $<$-increasing; so, in particular, it consists of pairwise distinct braids.*

Proof. By definition, every word in $\boldsymbol{w}_3^d$ is σ-positive. Hence, by Property A, it does not represent 1. □

As S_3^d consists of pairwise distinct divisors of Δ_3^d, Lemma 4.10 implies that every divisor of Δ_3^d occurs exactly once in S_3^d. Then, as S_3^d is $<$-increasing, it must be the increasing enumeration of $\mathrm{Div}(\Delta_3^d)$, and the proof of Proposition 4.7 is complete.

Remark 4.16. Once we know that S_3^d is the increasing enumeration of $\mathrm{Div}(\Delta_3^d)$ and that $\boldsymbol{w}_3^d$ is a σ-positive quotient sequence for S_3^d, we can count the 2-jumps in S_3^d and obtain the value of $h_2(\Delta_3^d)$ directly. This amounts to forgetting about all $\sigma_1^{\pm 1}$ in the construction of $\boldsymbol{w}_3^d$, and it is then fairly obvious that there only remains $2^d - 2$ times σ_2.

4D. ***Larger values of n.*** The same construction can be developed for $n = 4$ and beyond. The general scheme is to define S_4^d using an inductive rule

$$S_4^d = \theta_{4,0} S_3^d + S_4^{d,1} + \theta_{4,1} S_3^d + \cdots + \theta_{4,3d-1} S_3^d + S_4^{d,3d} + \theta_{4,3d} S_3^d,$$

where the intermediate factor $S_4^{d,p}$ is constructed by concatenating and translating convenient fragments of S_4^{d-1}. Owing to the inductive rule (3-6) satisfied by the number of elements $h_1(\Delta_4^d)$ of $\mathrm{Div}(\Delta_4^d)$, we can expect the generic entry of S_4^{d-1} to be repeated six times in S_4^d, but with some entries from S_4^{d-2} repeated three times only. After completing the inductive definition of S_4^d, showing that the sequence is $<$-increasing and counting its entries should be easy. As we have no complete description so far, we leave the question open here.

4E. ***A new construction for the linear ordering of B_3.*** In pursuing the approach described above, we were interested in connecting the Garside structure of B_n with its linear ordering. In the process, we found something more: a new, independent construction of the braid ordering, at least for B_3, which is currently the only completed case.

As recalled in the introduction, the existence of the linear ordering of braids relies on two properties of braids, namely Property A and Property C. These properties have received a number of independent proofs [Dehornoy et al. 2002]. In particular, Property A has now a very short proof based on Dynnikov's coordinization for singular triangulations of a punctured disk [Dehornoy et al. 2002, Chapter 9]. As for Property C, no really simple proof exists so far. Even without the initial argument involving self-distributive algebra, the remaining arguments—the combinatorial proofs based on handle reduction or on Burckel's uniform tree approach, or the geometric proofs based on standardization of curve diagrams—all require some care. For now, it seems that the optimal proof of Property C is forthcoming.

Here is a direct application of our construction of the sequence S_3^d:

Proposition 4.17. *Property C holds for* B_3; *that is, every nontrivial* 3*-braid admits a σ-positive or a σ-negative expression.*

New proof. We take as an hypothesis that Property A is true, so that the relation $<$ is a partial ordering, but we do not assume that $<$ is linear. As every braid in B_3 is the quotient of two positive braids in B_3^+, proving Property C for B_3 amounts to proving that, if x, y are arbitrary elements of B_3^+, then the quotient $x^{-1}y$ admits a σ-positive or a σ-negative expression.

Now the construction of S_3^d is self-contained, as is that of $\boldsymbol{w}_3^d$. Then, by construction, every word in $\boldsymbol{w}_3^d$ is σ-positive. As any concatenation of σ-positive words is σ-positive, it follows that, if x, y are any braids occurring in $\bigcup_d S_3^d$, then the quotient $x^{-1}y$ admits a σ-positive or a σ-negative expression, according to whether x occurs before or after y in S_3^d. To conclude Property C is true, it remains to check that each positive 3-braid occurs in $\bigcup_d S_3^d$. Because every entry of S_3^d belongs to $\mathrm{Div}(\Delta_3^d)$, this is equivalent to proving that each divisor of Δ_3^d occurs in S_3^d. Property A guarantees that the entries of S_3^d are pairwise distinct (Corollary 4.15), so it suffices to compare the length of S_3^d with the cardinality of $\mathrm{Div}(\Delta_3^d)$, and this is what we made in Lemma 4.10. □

The construction of S_3^d gives more. The approach developed by S. Burckel [1997] introduces a convenient notion of normal braid words such that every positive braid admits exactly one normal expression. For 3-strand braids, the definition is as follows. Every positive 3-strand braid word w can be written as an alternating product of blocks σ_1^e and σ_2^e. Then we define the *code* of w to be the sequence of the sizes of these blocks. To avoid ambiguity, we consider the last block to be a block of σ_1's, that is, we decide that the code of σ_1 is (1), while the code of σ_2 is $(1, 0)$. For instance, the code of $\sigma_2^2\sigma_1^3\sigma_2^5$ is $(2, 3, 5, 0)$.

Definition 4.18. A positive 3 strand braid word w is said to be *normal in the sense of Burckel* if its code has the form $(e_1, \dots, e_\ell)$ with $e_k \geqslant 2$ for $2 \leqslant k \leqslant \ell - 2$.

Burckel [1997] shows that every positive 3-braid admits a unique normal expression and, moreover, that $x < y$ holds if and only if the normal form of x is `ShortLex`-smaller than the normal form of y, where `ShortLex` refers to the variant of the lexicographic ordering of sequences in which the length is given priority: $(e_1, \dots, e_\ell) <_{\texttt{ShortLex}} (e'_1, \dots, e'_{\ell'})$ always holds for $\ell < \ell'$ and, when $\ell = \ell'$, it holds when $(e_1, \dots, e_\ell)$ is lexicographically smaller than $(e'_1, \dots, e'_{\ell'})$. Burckel's method defines an iterative reduction process on nonnormal braid words. Our current approach provides for a simpler method. First, a direct inspection shows:

Lemma 4.19. *Let $\underline{S}_3^d$ be the sequence of braid words defined by the inductive rule* (4-5). *Then $\underline{S}_3^d$ consists of words that are normal in the sense of Burckel.*

Then, by construction, every braid in S_3^d is represented by a word of $\underline{S}_3^d$. As every positive 3-braid occurs in $\bigcup S_3^d$, we immediately deduce:

Proposition 4.20. *Every positive* 3*-braid admits an expression that is normal in the sense of Burckel.*

This in turn enables us to obtain a simple proof for the following deep, and so far not very well understood, result due to Laver [1996] and to Burckel [1997] for the ordinal type:

Corollary 4.21. *The restriction of $<$ to B_3^+ is a well-ordering of ordinal type ω^ω.*

Proof. The ShortLex ordering of sequences of nonnegative integers is a well-ordering of ordinal type ω^ω, so its restriction to codes of normal words in the sense of Burckel is a well-ordering as well. The type of the latter cannot be less than ω^ω, as one can easily exhibit an increasing sequence of length ω^ω. □

Burckel's approach extends to all braid monoids B_n^+. Burckel introduces a convenient notion of a normal word, but the associated reduction process is very intricate. Hopefully, the above approach will provide a much simpler approach to completing the construction of the sequences S_4^d and, more generally, S_n^d. In particular, once the correct definition is given, all subsequent proofs should reduce to easy inductions.

Acknowledgement

The author thanks the PJM production editors, Matt Cargo and Silvio Levy, for their suggestions.

References

[Birman 1974] J. S. Birman, *Braids, links, and mapping class groups*, Princeton University Press, Princeton, N.J., 1974. Annals of Mathematics Studies, No. 82. MR 51 #11477

[Burckel 1997] S. Burckel, "The wellordering on positive braids", *J. Pure Appl. Algebra* **120**:1 (1997), 1–17. MR 98h:20062 Zbl 0958.20032

[Burckel 1999] S. Burckel, "Computation of the ordinal of braids", *Order* **16**:3 (1999), 291–304. MR 2001i:20075 Zbl 0980.20026

[Burckel 2001] S. Burckel, "Syntactical methods for braids of three strands", *J. Symbolic Comput.* **31**:5 (2001), 557–564. MR 2002b:20051 Zbl 0990.20022

[Carlitz et al. 1976] L. Carlitz, R. Scoville, and T. Vaughan, "Enumeration of pairs of permutations", *Discrete Math.* **14**:3 (1976), 215–239. MR 53 #156 Zbl 0322.05008

[Charney 1995] R. Charney, "Geodesic automation and growth functions for Artin groups of finite type", *Math. Ann.* **301**:2 (1995), 307–324. MR 95k:20055 Zbl 0813.20042

[Dehornoy 1994] P. Dehornoy, "Braid groups and left distributive operations", *Trans. Amer. Math. Soc.* **345**:1 (1994), 115–150. MR 95a:08003 Zbl 0837.20048

[Dehornoy 1997] P. Dehornoy, "A fast method for comparing braids", *Adv. Math.* **125**:2 (1997), 200–235. MR 98b:20060 Zbl 0882.20021

[Dehornoy 1999] P. Dehornoy, "Strange questions about braids", *J. Knot Theory Ramifications* **8**:5 (1999), 589–620. MR 2000d:20052 Zbl 0933.20024

[Dehornoy 2002] P. Dehornoy, "Groupes de Garside", *Ann. Sci. École Norm. Sup.* (4) **35**:2 (2002), 267–306. MR 2003f:20068 Zbl 1017.20031

[Dehornoy 2007] P. Dehornoy, "Combinatorics of normal sequences of braids", *J. Combin. Theory Ser. A* **114**:3 (2007), 389–409. MR MR2310741 Zbl 1116.05006

[Dehornoy et al. 2002] P. Dehornoy, I. Dynnikov, D. Rolfsen, and B. Wiest, *Why are braids orderable?*, Panoramas et Synthèses **14**, Société Mathématique de France, Paris, 2002. MR 2004e:20062 Zbl 1048.20021

[El-Rifai and Morton 1994] E. A. El-Rifai and H. R. Morton, "Algorithms for positive braids", *Quart. J. Math. Oxford Ser.* (2) **45**:180 (1994), 479–497. MR 96b:20052 Zbl 0839.20051

[Epstein et al. 1992] D. B. A. Epstein, J. W. Cannon, D. F. Holt, S. V. F. Levy, M. S. Paterson, and W. P. Thurston, *Word processing in groups*, Jones and Bartlett, Boston, 1992. MR 93i:20036 Zbl 0764.20017

[Garside 1969] F. A. Garside, "The braid group and other groups", *Quart. J. Math. Oxford Ser.* (2) **20** (1969), 235–254. MR 40 #2051 Zbl 0194.03303

[Larue 1994] D. Larue, *Left-distributive and left-distributive idempotent algebras*, Ph.D. thesis, University of Colorado, Boulder, 1994.

[Laver 1996] R. Laver, "Braid group actions on left distributive structures, and well orderings in the braid groups", *J. Pure Appl. Algebra* **108**:1 (1996), 81–98. MR 97e:20061 Zbl 0859.20029

[McCammond 2005] J. McCammond, "An introduction to Garside structures", Preprint, 2005, Available at http://www.math.ucsb.edu/~mccammon/papers/intro-garside.pdf.

[Solomon 1976] L. Solomon, "A Mackey formula in the group ring of a Coxeter group", *J. Algebra* **41**:2 (1976), 255–264. MR 56 #3104 Zbl 0355.20007

Received April 6, 2006.

PATRICK DEHORNOY
LABORATOIRE DE MATHÉMATIQUES NICOLAS ORESME UMR 6139
UNIVERSITÉ DE CAEN
14032 CAEN
FRANCE

dehornoy@math.unicaen.fr
www.math.unicaen.fr/~dehornoy

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

EXISTENCE OF INFINITELY MANY EQUILIBRIUM CONFIGURATIONS OF A LIQUID CRYSTAL SYSTEM PRESCRIBING THE SAME NONCONSTANT BOUNDARY VALUE

MIN-CHUN HONG

In 1986, Hardt, Kinderlehrer and Lin established the existence and partial regularity of minimizers of the liquid crystal energy with the Oseen–Frank density. Motivated by the earlier results of Bethuel–Brezis–Coron and Riviere on harmonic maps, we prove the existence of infinitely many equilibrium configurations of the liquid crystal energy prescribing the same nonconstant boundary data.

1. Introduction

Let $\Omega \subset \mathbb{R}^3$ be a domain with smooth boundary $\partial\Omega$, and let $\gamma : \partial\Omega \to S^2$ be smooth boundary data. The equilibrium configuration of a liquid crystal is described by a unit vector field u on Ω. For any map $u \in H^1_\gamma(\Omega, S^2)$ with $\gamma : \partial\Omega \to S^2$, the integral

$$\Phi(\gamma) = \tfrac{1}{2} \int_\Omega \left[\operatorname{tr}(\nabla u)^2 - (\operatorname{div} u)^2\right] dx$$

depends only on γ [Hardt et al. 1986]. According to the Ericksen–Leslie theory [Giaquinta et al. 1998], the Oseen–Frank bulk energy of a configuration $u \in H^1(\Omega, S^2)$ can be reduced to

$$E(u, \Omega) = \int_\Omega W(u, \nabla u)\, dx, \tag{1–1}$$

where $W(u, \nabla u)$ is the Oseen–Frank density

$$\begin{aligned} W(u, \nabla u) = {}& \alpha|\nabla u|^2 \\ &+ (k_1 - \alpha)(\operatorname{div} u)^2 + (k_2 - \alpha)(u \cdot \operatorname{curl} u)^2 + (k_3 - \alpha)\,|u \times \operatorname{curl} u|^2, \end{aligned} \tag{1–2}$$

with constants $k_1 > 0$, $k_2 > 0$, $k_3 > 0$, and $\alpha = \min\{k_1, k_2, k_3\}$.

MSC2000: primary 35J50; secondary 35Q99.
Keywords: liquid crystal, equilibrium configurations, harmonic maps.

For the above density $W(u, \nabla u)$, we set $V(u, p) = W(u, p) - \alpha |p|^2$ with $p = \nabla u$. As in [Hardt et al. 1986], the equilibrium system associated to E is of the form

$$-\operatorname{div}\big(W_p(u, \nabla u) - u \otimes u V_p(u, \nabla u)\big) + Y(u, \nabla u) = 0 \quad \text{in } \Omega, \tag{1–3}$$

where

$$Y(u, \nabla u) = (I - u \otimes u)\, W_u(u, \nabla u) - \nabla u \big(u V_p(u, \nabla u)\big) - \big(W_p(u, \nabla u) \cdot \nabla u\big)\, u$$

satisfies $|Y(u, p)| \le C |p|^2$ for all $p = (p_i^j)_{3\times 3}$ with $p_i^j \in \mathbb{R}$.

A static equilibrium configuration u corresponds to an extremal of the functional (1–1) in $H^1(\Omega, S^2)$, that is, $u \in H^1(\Omega, S^2)$ is a weak solution of system (1–3).

In a special case $k_1 = k_2 = k_3$, the equilibrium system (1–3) is

$$\triangle u + |\nabla u|^2 u = 0 \quad \text{in } \Omega,$$

which is the equation for harmonic maps. When $k_1 = k_2 = k_3 = 1$, Bethuel et al. [1990] first proved the existence of infinitely many harmonic maps for some special boundary values γ. Rivière [1995] proved the existence of infinite many harmonic maps for all nonconstant boundary values. See further generalizations to higher dimensions in [Isobe 1995] and [Pakzad 2001].

In general, (1–3) is not always elliptic for every choice of the constants k_i, and so the system (1–3) is much more complicated than the harmonic map equation. Hardt et al. [1986], in a fundamental paper, proved the existence and partial regularity of a minimizer u, which is a weak solution of system (1–3) that gives the energy E in $H^1_\gamma(\Omega, S^2)$ given boundary data $\gamma : \partial\Omega \to S^2$. One questions whether one can prove there exist infinitely many weak solutions of the liquid crystal system (1–3) prescribing the same boundary data. Here, we prove the existence of infinitely many equilibrium configurations prescribing the same nonconstant boundary data γ in:

Theorem 1.1. *Let $\gamma : \partial\Omega \to S^2$ be a nonconstant smooth map. Assume that the constants k_1, k_2 and k_3 in* (1–2) *satisfy $|k_1 - k_2| \le \min\{k_1, k_2\} 4(1 - \ln 2)/\ln 2$. Then there exist infinitely many stable weak solutions of system* (1–3) *in $H^1(\Omega, S^2)$ with the same boundary value γ.*

The key to proving Theorem 1.1 is generalizing the idea of Rivière [1995] to the liquid crystal energy. Riviere's idea relies on constructing dipoles and the relaxed energy of the Dirichlet energy in [Brezis et al. 1986] and [Bethuel et al. 1990]. More precisely, Riviere inserts a dipole into nonconstant maps and finds a way to confine the energy to strictly less than 8π times the length of the dipole.

In this paper, we extend a key result of Riviere to the liquid crystal:

Theorem 1.2. *Assume that the constants k_1, k_2 and k_3 in* (1–2) *satisfy the condition* $|k_1 - k_2| \le \min\{k_1, k_2\} 4(1 - \ln 2)/\ln 2$. *Let $u \in H^1(\Omega, S^2)$ be a nonconstant and smooth map, and let x_0 be a point inside Ω such that $\nabla u(x_0) \neq 0$. For any $\rho > 0$ with $B_\rho(x_0) \subset \Omega$, there exist two points $P, N \in B_\rho(x_0)$ with middle point x_0 and $|P - N| = 2\sigma \le 2\rho$ and a map v in $H^1(\Omega, S^2) \cap C^{0,1}(\Omega \backslash \{P, N\}, S^2)$ such that*

$$v = u \quad \textit{in } \Omega \backslash B_\rho(x_0), \qquad \deg(v, P) = -\deg(v, N) = +1,$$

and

$$\int_\Omega W(v(x), \nabla v(x))\, dx < \int_\Omega W(u(x), \nabla u(x))\, dx + 8\pi \Gamma(k_1, k_2, k_3)\, |P - N| .$$

For the proof of Theorem 1.2, thanks to Giaquinta et al. [1990], the dipoles and relaxed functional of the liquid crystal energy E was given: For any $u \in H^1(\Omega, S^2)$, the vector field $D(u)$ is defined by

$$D(u) = (u \cdot u_{x_2} \wedge u_{x_3},\ u \cdot u_{x_3} \wedge u_{x_1},\ u \cdot u_{x_1} \wedge u_{x_2}).$$

Given $u^0 \in H^1_\gamma(\Omega, S^2)$, we set

$$L(u, u^0) := \frac{1}{4\pi} \sup_{\substack{\xi : \Omega \to \mathbb{R}, \\ \|\nabla \xi\|_{L^\infty} \le 1}} \int_\Omega \big[D(u) - D(u^0)\big] \cdot \nabla \xi \, dx$$

for maps $u \in H^1_\gamma(\Omega, S^2)$. The relaxed functional F_{u^0} of the liquid crystal energy E is given by

$$F_{u^0}(u) = \int_\Omega W(u(x), \nabla u(x))\, dx + 8\pi \Gamma(k_1, k_2, k_3) L(u, u^0),$$

where

$$\Gamma(k_1, k_2, k_3) = \sqrt{k k_3} \int_0^1 \sqrt{1 + (k/k_3 - 1)\, s^2}\, ds \ \ge \alpha, \quad k = \min\{k_1, k_2\}.$$

More precisely, $F_{u^0}(u)$ is lower semicontinuous in the weak H^1-topology and any minimizer of $F_{u^0}(u)$ in H^1_γ is also a weak solution of system (1–3) prescribing the boundary value γ. One key to proving Theorem 1.2 is obtaining new estimates on the irrotational and solenoidal dipole in [Giaquinta et al. 1990] for inserting a small dipole into nonconstant map. More precisely, for $k_2 \ge k_1$, one finds the irrotational map u (with $u \cdot \operatorname{curl} u = 0$) on $\mathbb{R}^2 \times (0, l)$ of the form

$$u(x) = \left(g(r) \frac{x_1}{r},\, g(r) \frac{x_2}{r},\, \operatorname{sign}(1 - r) \sqrt{1 - g^2(r)} \right), \quad r = \sqrt{]x_1^2 + x_2^2}$$

such that $E(u) = 8\pi l \Gamma(k_1, k_2, k_3)$, where $g \in C([0, \infty); [0, 1])$, $g'(r) > 0$ on $(0, 1)$, and $g'(r) < 0$ on $(1, \infty)$. Similarly, for $k_2 \le k_1$, one finds the solenoidal map (with

$\operatorname{div} u = 0$) on $\mathbb{R}^2 \times (0, l)$ of the form

$$u(x) = \left(g(r)\frac{x_2}{r}, -g(r)\frac{x_1}{r}, \operatorname{sign}(1-r)\sqrt{1-g^2(r)} \right)$$

such that $E(u) = 8\pi l \Gamma(k_1, k_2, k_3)$, where $g \in C([0, \infty); [0, 1])$ and $g'(r) > 0$ on $(0, 1)$, and $g'(r) < 0$ on $(1, \infty)$. In this paper, we derive new estimates on $g(r)$ and also prove that $g'(0)$ exists, is positive, and is bounded by a constant depending on k_1, k_2 and k_3.

The second key step is to improve the method in [Rivière 1995] (also [Brezis and Coron 1983]) of inserting a small dipole into a nonconstant map. During the proof, it is important that $g'(0)$ is positive and bounded. Due to the differing constants k_1, k_2, and k_3, the liquid crystal is more complicated and involved than the harmonic maps.

Theorem 1.1 is a consequence of Theorem 1.2. Rivière [1995] also used Theorem 1.2 to construct weak harmonic maps having singularities almost everywhere in Ω. With Theorem 1.2, we conjecture that one can construct a weak solution of system (1–3) having singularities almost everywhere in Ω for different constants k_1, k_2, and k_3.

In the last part, we deal with the partial regularity of weak solutions of system (1–3). The partial regularity of weak solutions of elliptic systems and weakly harmonic maps has been of great interest (for example [Giaquinta 1983; Giaquinta et al. 1998]). For the liquid crystal, Hardt et al. [1986; 1988], in fundamental papers, proved the partial regularity of minimizers of the liquid crystal energy E. Here, we investigate the partial regularity of the weak solutions that minimize a modified relaxed functional of the liquid crystal energy E.

For a parameter $\lambda \in [0, 1]$, as in [Bethuel and Brézis 1991], we consider the modified λ-energy

$$E_\lambda(u) := E(u) + \lambda 8\pi \Gamma(k_1, k_2, k_3) L(u, u^0) \tag{1–4}$$

for a map $u \in H^1_\gamma(\Omega, S^2)$. It follows from [Bethuel et al. 1990] and [Giaquinta et al. 1990] that there exists a minimizer u_λ of E_λ in $H^1_\gamma(\Omega, S^2)$, and u_λ is a weak solution of (1–3). The author in [Hong 2004] proved the partial regularity of minimizers u_λ for $0 \le \lambda < \lambda_0 = \alpha/\Gamma(k_1, k_2, k_3)$ with $\Gamma(k_1, k_2, k_3) \ge \alpha$. It was not clear then whether one can establish the partial regularity of minimizers u_λ of (1–4) for $\lambda \in [\lambda_0, 1]$. Now we make progress with:

Theorem 1.3. *For any parameter λ with $0 \le \lambda < 1$, let u_λ be a minimizer of E_λ in $H^1_\gamma(\Omega, S^2)$. Then u_λ is smooth in a set $\Omega_0 \subset \bar{\Omega}$ and $\mathcal{H}^\beta(\bar{\Omega} \backslash \Omega_0) = 0$ for some positive $\beta < 1$, where $\mathcal{H}^\beta$ is the Hausdorff measure.*

The paper is organized as follows. In Section 2, we derive some new estimates for the irrotational dipole and the solenoidal dipole. In Section 3, we prove Theorems 1.1 and 1.2 for $k_2 \geq k_1$. In Section 4, we prove those theorems for $k_1 > k_2$. In Section 5, we complete a proof of Theorem 1.3.

2. Improving estimates for irrotational dipole and solenoidal dipole

Proposition 2.1 [Giaquinta et al. 1990]. *There exists a C^∞ function $\tilde{u}(x)$ from $\mathbb{R}^2$ into $S^2 \subset \mathbb{R}^3$ such that*

(i) *$\tilde{u} = q$ at infinity, where q is the south pole of $S^2 \subset \mathbb{R}^3$;*

(ii) *$\tilde{u}$, seen as a map from S^2 into S^2, has degree* 1;

(iii) *if $\Omega := \mathbb{R}^2 \times (0, l)$ and $u_0 : \Omega \to S^2$ is defined as*

$$u_0(x_1, x_2, x_3) = \tilde{u}(x_1, x_2),$$

we have

$$E\left(u_0, \mathbb{R}^2 \times (0, l)\right) = 8\pi l \Gamma(k_1, k_2, k_3). \tag{2–1}$$

Now we will improve Proposition 2.1 so we can apply it to prove Theorem 1.2. Consider the dipole

$$T_0 = G_q + L \times [\![S^2]\!], \quad \text{where } L = [\![(0, 0, x_3) : 0 < x_3 \leq l]\!],$$

that is, $P = (0, 0, 0)$ and $N = (0, 0, l)$ with $l > 0$, where G_q is the current by the graph of the constant function q.

From [Giaquinta et al. 1990], we have

$$E\left(T_0, \mathbb{R}^2 \times (0, l)\right) = 2l \int_{S^2} \sqrt{k^2 n_3^2 + k k_3 (1 - n_3^2)}\, dH^2(n),$$

where $n = (n_1, n_2, n_3) \in S^2$. Then

$$\begin{aligned} E\left(u_0, \mathbb{R}^2 \times (0, l)\right) = E\left(T_0, \mathbb{R}^2 \times (0, l)\right) &= 8\pi k l \int_0^1 \sqrt{k_3/k + (1 - k_3/k) z^2}\, dz \\ &= 8\pi k l \int_0^1 \frac{\sqrt{1 - \beta y^2}}{\sqrt{1 - y^2}}\, y\, dy, \end{aligned}$$

where $\beta = 1 - k_3/k$.

The irrotational dipole. Assume $k_2 \geq k_1 = k$. We consider the all maps $u = (u_1, u_2, u_3) : \mathbb{R}^3 \to S^2$ of form

$$u_1(x) = g(r)\frac{x_1}{r}, \quad u_2 = g(r)\frac{x_2}{r}, \quad u_3(x) = \begin{cases} \sqrt{1 - g^2(r)}, & \text{for } 0 \leq r \leq 1, \\ -\sqrt{1 - g^2(r)}, & \text{for } r > 1, \end{cases}$$

where $g:[0,+\infty)\to[0,1]$ is continuous and satisfies $g(0)=0$, $g(1)=1$, $g(r)\to 0$ as $r\to+\infty$, $g'(r)>0$ on $(0,1)$, and $g'(r)<0$ on $(1,+\infty)$.

By a standard calculation, we have

$$(\operatorname{div} u)^2=(g')^2+\frac{g^2}{r^2}+2gg'\frac{1}{r},$$
$$|u\times\operatorname{curl} u|^2=|\operatorname{curl} u|^2=|\nabla u_3|^2=\frac{g^2(g')^2}{1-g^2},$$

and

$$u\cdot\operatorname{curl} u=0,\quad |\nabla u_1|^2+|\nabla u_2|^2=(g')^2+\frac{g^2}{r^2}.$$

The Oseen–Frank density becomes

$$W(u,\nabla u)=k_1\left[\frac{g^2}{r^2}+g'^2\frac{(1-\beta)g^2}{1-g^2}\right]+2(k_1-\alpha)gg'\frac{1}{r}, \tag{2–2}$$

where $\beta=1-k_3/k$. Then

$$E\left(u,\mathbb{R}^2\times(0,l)\right)=2\pi kl\int_0^\infty\left[\frac{g^2}{r^2}+g'^2\frac{1-\beta g^2}{1-g^2}\right]r\,dr.$$

So

$$E(T_0)=4\pi kl\int_0^\infty\frac{g}{r}g'\frac{\sqrt{1-\beta g^2}}{\sqrt{1-g^2}}r\,dr\le 2\pi kl\int_0^\infty\left[\frac{g^2}{r^2}+g'^2\frac{1-\beta g^2}{1-g^2}\right]r\,dr=E(u),$$

with equality if and only if

$$g'=\begin{cases}-\dfrac{g}{r}\dfrac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}}, & \text{for } r\ge 1,\\[2ex] \dfrac{g}{r}\dfrac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}}, & \text{for } 0\le r\le 1.\end{cases} \tag{2–3}$$

First we consider the case $0\le r\le 1$. We will prove that there is a solution of the equation

$$g'=\frac{g}{r}\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}} \tag{2–4}$$

with $g(0)=0$ and $g(1)=1$. Moreover, $g'(0)$ exists and is positive and bounded.

Case I: $k\ge k_3$.

Since $0 \le \beta \le 1$, we have $1-g^2 \le 1-\beta g^2 \le 1$. Then

$$\frac{g}{r} \ge \frac{g}{r}\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}} \ge \frac{g}{r}\sqrt{1-g^2}.$$

We consider two auxiliary equations:

$$g_1' = \frac{g_1}{r},$$
$$g_2' = \frac{g_2}{r}\sqrt{1-g_2^2}.$$

It is easy to see that $g_1(r) = r$ solves the first with $g_1(0) = 0$ and $g_1(1) = 1$ and $g_2(r) = 2r/1+r^2$ solves the second with $g_2(0) = 0$ and $g_2(1) = 1$.

By the comparison theorem, there is a solution g of (2–4) such that

$$r \le g \le \frac{2r}{1+r^2}, \quad g(0) = 0, \quad g(1) = 1. \tag{2–5}$$

Using Equation (2–3), we have

$$\begin{aligned} g''(r) &= \frac{g'}{r}\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}} - \frac{g}{r^2}\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}} - \frac{g}{r}\frac{gg'}{\sqrt{1-g^2}\sqrt{1-\beta g^2}} + \frac{g}{r}\frac{\sqrt{1-g^2}\beta gg'}{(1-\beta g^2)^{3/2}} \\ &= \frac{g}{r^2}\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}}\left(\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}} - 1\right) - \frac{g^2g'}{r}\frac{1-\beta}{(1-g^2)^{1/2}(1-\beta g^2)^{2/3}} \le 0. \end{aligned}$$

Therefore $g'(0) = \lim_{r\to 0} g'(r)$ exists and is finite because $g_1'(0) = 1$ and $g_2'(0) = 2$. More precisely, we know

$$1 \le g'(0) \le 2.$$

Case II: $k_3 \ge k$.

Since $\beta \le 0$, we have $1-\beta \ge 1-\beta g^2 \ge 1$. Then

$$\frac{1}{\sqrt{1-\beta}}\frac{g}{r}\sqrt{1-g^2} \le \frac{g}{r}\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}} \le \frac{g}{r}\sqrt{1-g^2}.$$

Then we consider two auxiliary equations:

$$g_1' = \frac{1}{\sqrt{1-\beta}}\frac{g_1}{r}\sqrt{1-g_1^2},$$
$$g_2' = \frac{g_2}{r}\sqrt{1-g_2^2}.$$

Note that $g_1(r) = 2r^c/1+r^{2c}$ with $c = 1/\sqrt{1-\beta}$ solves the first with $g_1(0) = 0$ and $g_1(1) = 1$ and $g_2(r) = 2r/1+r^2$ solves the second with $g_2(0) = 0$ and $g_2(1) = 1$.

By the comparison theorem, there is a solution g of the equation with $g(0)=0$ and $g(1)=1$ such that

$$\frac{2r}{1+r^2} \le g \le \frac{2r^c}{1+r^{2c}}.$$

Then

$$\frac{1}{\sqrt{1-\beta g^2}} \ge \frac{1}{1-\beta g^2} = 1+\frac{\beta g^2}{1-\beta g^2} \ge 1-|\beta|\, g^2 \ge 1-4\,|\beta|\,\frac{r^{2c}}{(1+r^{2c})^2}.$$

Note that as $r\to 0$, $g_1'(r)\to +\infty$. So we need to consider the auxiliary equation

$$g_3' = \left(1-4\,|\beta|\,\frac{r^{2c}}{(1+r^{2c})^2}\right)\frac{g_3}{r}\sqrt{1-g_3^2} \tag{2–6}$$

with $g_3(0)=0$ and $g_3(1)=1$. The solution to Equation (2–6) is

$$g_3(r) = 2r\exp\left(\frac{|\beta|}{c}\frac{1-r^{2c}}{1+r^{2c}}\right)\Big/\left(1+r^2\exp\left(\frac{2|\beta|}{c}\frac{1-r^{2c}}{1+r^{2c}}\right)\right).$$

By the comparison theory, we have

$$g_2(r)\le g(r)\le g_3(r) = 2r\exp\left(\frac{|\beta|}{c}\frac{1-r^{2c}}{1+r^{2c}}\right)\Big/\left(1+r^2\exp\left(\frac{2|\beta|}{c}\frac{1-r^{2c}}{1+r^{2c}}\right)\right) \tag{2–7}$$

It is easy to see that $g'(0)$ exists and is finite because $g_2'(0)=2$ and $g_3'(0)=2e^{|\beta|/c}$. More precisely, we have

$$2\le g'(0)\le 2e^{|\beta|/c}.$$

In both Cases I and II, there is a solution with $g(0)=0$, $g(1)=1$ and with $g'(0)$ positive and finite.

If $r\ge 1$, take $h(r)=g(r^{-1})$, where $g(r)$ solves Equation (2–4) with $g(0)=0$ and $g(1)=1$. Using Equation (2–4), we have

$$h'(r) = g'\left(r^{-1}\right)\left(-r^{-2}\right) = \frac{g\left(r^{-1}\right)}{r^{-1}}\frac{\sqrt{1-g^2}}{\sqrt{1-\beta g^2}}\left(-r^{-2}\right) = -\frac{h}{r}\frac{\sqrt{1-h^2}}{\sqrt{1-\beta h^2}}.$$

Then

$$\tilde g(r) = \begin{cases} g(r), & \text{for } 0\le r\le 1,\\ h(r), & \text{for } r>1,\end{cases}$$

is the required solution of Equation (2–3) with $\tilde g(r)=\tilde g(r^{-1})$.

The solenoidal dipole. Assume $k_2\ge k_1=k$. Consider all maps $u=(u_1,u_2,u_3):\Omega\to S^2$ of the form

$$u_1(x)=g(r)\frac{x_2}{r},\quad u_2=-g(r)\frac{x_1}{r},\quad u_3(x)=\begin{cases}\sqrt{1-g^2(r)}, & \text{for } 0\le r\le 1\\ -\sqrt{1-g^2(r)}, & \text{for } r>1,\end{cases}$$

where $g:[0,+\infty)\to[0,1]$ is continuous and satisfies $g(0)=0$, $g(1)=1$, $g(r)\to 0$ as $r\to+\infty$, $g'(r)>0$ on $(0,1)$, and $g'(r)<0$ on $(1,+\infty)$. Then we have

$$E\left(u,\mathbb{R}^2\times(0,l)\right)=2\pi kl\int_0^\infty\left[\frac{g^2}{r^2}(1-\beta g^2)+\frac{(g')^2}{1-g^2}\right]r\,dr$$

such that $E\left(T_0,\mathbb{R}^2\times(0,l)\right)\le E(u)$, with equality if and only if

$$\frac{dg}{dr}=\operatorname{sign}(1-r)\,\frac{g}{r}\sqrt{1-g^2}\sqrt{1-\beta g^2}, \tag{2–8}$$

This equation has a solution $g(r)$ such that $g:[0,+\infty)\to[0,1]$ is continuous and satisfies $g(0)=0$, $g(1)=1$, $g(r)\to 0$ as $r\to+\infty$, $g'(r)>0$ on $(0,1)$, and $g'(r)<0$ on $(1,+\infty)$. Moreover, $g'(0)$ exists and is positive and bounded.

3. Proof of Theorems 1.1 and 1.2 for $k_2\ge k_1$

3.1. ***The construction of u^δ for $k_2\ge k_1$.*** We assume Theorem 1.2 that $\nabla u(x_0)\neq 0$. Without losing generality, we also assume $x_0=0$. Note that

$$W(Qu,Q\nabla uQ^T)=W(u,\nabla u),\quad \text{for all } Q\in \mathrm{O}(3).$$

After a rotation Q on both $x,u\in\mathbb{R}^3$, we can choose an orthonormal basis $\{I,J,K\}$ of $\mathbb{R}^3$ for both $x,u\in\mathbb{R}^3$ as in [Brezis and Coron 1983] such that $u(0,0,0)=K$. $u_{x_1}(0,0,0)\cdot u_{x_2}(0,0,0)=0$ and $u_{x_1}(0,0,0)\neq 0$.

Without loss of generality, we may choose

$$K(x_3)=u(0,0,x_3),\qquad I(x_3)=\frac{u_{x_1}(0,0,x_3)}{|u_{x_1}(0,0,x_3)|}.$$

to form a basis $\{I(x_3),J(x_3),K(x_3)\}$ of $\mathbb{R}^3$ depending on x_3. We write

$$u=\hat u_1 I(x_3)+\hat u_2 J(x_3)+\hat u_3 K(x_3)$$

with $\hat u_1(0,0,x_3)=\hat u_2(0,0,x_3)=0$ and $\hat u_3(0,0,x_3)=1$. More precisely, there are two numbers $a>0$ and $b\ge 0$ — this is true after a rotation in $\mathbb{R}^3$: after a rotation in the subspace $\mathbb{R}^2$ of $\mathbb{R}^3$, the conclusion of Theorem 1.2 does not change — such that

$$u_{x_1}(0,0,x_3)=(a+O(x_3))I(x_3),\quad u_{x_2}(0,0,x_3)=O(x_3)I(x_3)+(b+O(x_3))J(x_3).$$

We use polar coordinates for $(x_1,x_2)\in\mathbb{R}^2$, that is,

$$x_1=r\cos\theta,\quad x_2=r\sin\theta,$$

and consider the cylinder C^δ in $\mathbb{R}^3$ defined by

$$C^\delta=\left\{(x_1,x_2,x_3)\in\mathbb{R}^3\;\middle|\;0\le r\le\delta+\delta^2,\ -\delta-\delta^2\le x_3\le\delta+\delta^2\right\}.$$

As in [Rivière 1995], we construct a map

$$u^\delta = \hat{u}_1^\delta I(x_3) + \hat{u}_2^\delta J(x_3) + \hat{u}_3^\delta K(x_3)$$

such that

(i) $u^\delta = u$ outside C^δ.

(ii) Inside C^δ, for each $x_3 \in [-\delta + \delta^2, \delta - \delta^2]$ (that is, the subcylinder of C^δ), we construct u^δ in three different cases:

(a) If $r > 2\delta^2$, we set $u^\delta(x) = u(x)$;
(b) If $r < \delta^2$, we set

$$u^\delta (x_1, x_2, x_3) = \\ g(r/\lambda)\frac{x_1}{r} I(x_3) + g(r/\lambda)\frac{x_2}{r} J(x_3) + \operatorname{sign}(1-r)\sqrt{1-g^2(r/\lambda)}\, K(x_3).$$

(c) If $\delta^2 \le r \le 2\delta^2$, we set

$$u^\delta(x) = \\ (A_1 r + B_1)\, I(x_3) + (A_2 r + B_2)\, J(x_3) + \sqrt{1-(A_1 r+B_1)^2-(A_2 r+B_2)^2}\, K(x_3),$$

where A_1, A_2, B_1, B_2, depending only on θ, δ and x_3, are determined by

$$\begin{aligned} 2\delta^2 A_1 + B_1 &= \hat{u}_1(2\delta^2\cos\theta, 2\delta^2\sin\theta, x_3), \\ 2\delta^2 A_2 + B_2 &= \hat{u}_2(2\delta^2\cos\theta, 2\delta^2\sin\theta, x_3), \\ \delta^2 A_1 + B_1 &= g(\delta^2/\lambda)\cos\theta = g(\lambda/\delta^2)\cos\theta, \\ \delta^2 A_2 + B_2 &= g(\delta^2/\lambda)\sin\theta = g(\lambda/\delta^2)\sin\theta, \end{aligned} \tag{3–1}$$

and $g(r)$ is the solution of Equation (2–3) with $g(0) = 0$, $g'(0) > 0$, $g(1) = 1$, $g(r) = g(1/r)$, and $\lambda = c\delta^4$, where c will be determined later.

(iii) Inside C^δ, for each

$$x_3 \in [-\delta, -\delta + \delta^2] \cup [\delta, \delta - \delta^2],$$

we let $P = (0, 0, \delta)$ and $N = (0, 0, -\delta)$ in a small cylinder c_P^δ (or c_N^δ). The cylinder is centered at P (or N) with radius $2\delta^2$, length $2\delta^2$, and its axis along the x_3-axis. If we denote by Π^+ (or Π_-) the radial projection centered at P (or N) onto the boundary of c_P^δ (or c_N^δ), the transformed map u^δ is the composition of Π^+ (or Π_-) and the value of u^δ on this boundary.

3.2. ***The estimate of the energy of u^δ for $k_2 \geq k_1$.*** *Case 1.* The estimate of the energy of u^δ on the domain of $x_3 \in [-\delta + \delta^2, \delta - \delta^2]$ and $\delta^2 \le r \le 2\delta^2$.

Notice that $\hat{u}_i(0,0,x_3) = 0$ for $i = 1, 2$ and

$$\begin{aligned}
&\frac{\partial \hat{u}_1}{\partial x_1}(0,0,x_3) = a + O(x_3), && \frac{\partial \hat{u}_2}{\partial x_1}(0,0,x_3) = O(x_3),\\
&\frac{\partial \hat{u}_1}{\partial x_2}(0,0,x_3) = O(x_3), && \frac{\partial \hat{u}_2}{\partial x_2}(0,0,x_3) = b + O(x_3),\\
&g(c\delta^2) = c\delta^2 g'(0) + O(\delta^4)
\end{aligned}$$

Then we have

$$\begin{aligned}
2\delta^2 A_1 + B_1 &= 2a\delta^2 \cos\theta + O(\delta^3),\\
2\delta^2 A_2 + B_2 &= 2b\delta^2 \sin\theta + O(\delta^3),\\
\delta^2 A_1 + B_1 &= g'(0)c\delta^2 \cos\theta + O(\delta^4),\\
\delta^2 A_2 + B_2 &= g'(0)c\delta^2 \sin\theta + O(\delta^4).
\end{aligned}$$

Solving these equations, we have

$$\begin{aligned}
A_1 &= (2a - cg'(0))\cos\theta + O(\delta),\\
A_2 &= (2b - cg'(0))\sin\theta + O(\delta),\\
B_1 &= 2\delta^2(g'(0)c - a)\cos\theta + O(\delta^3),\\
B_2 &= 2\delta^2(g'(0)c - b)\sin\theta + O(\delta^3).
\end{aligned} \tag{3–2}$$

In a way similar to (3–2), it follows from (3–1) that

$$\begin{aligned}
\frac{\partial A_1}{\partial\theta} &= -(2a - cg'(0))\sin\theta + O(\delta),\\
\frac{\partial A_2}{\partial\theta} &= (2b - cg'(0))\cos\theta + O(\delta),\\
\frac{\partial B_1}{\partial\theta} &= -2\delta^2(cg'(0) - a)\sin\theta + O(\delta^3),\\
\frac{\partial B_2}{\partial\theta} &= 2\delta^2(cg'(0) - b)\cos\theta + O(\delta^3).
\end{aligned} \tag{3–3}$$

In polar coordinates, we know

$$\frac{\partial\theta}{\partial x_1} = -\frac{\sin\theta}{r}, \quad \frac{\partial\theta}{\partial x_2} = \frac{\cos\theta}{r}, \quad \frac{\partial r}{\partial x_1} = \cos\theta, \quad \frac{\partial r}{\partial x_2} = \sin\theta.$$

Using $\hat{u}_1(0,0,x_3) = \hat{u}_2(0,0,x_3) = 0$ in (3–1), we obtain, for $\delta^2 \le r \le 2\delta^2$,

$$\frac{\partial A_1}{\partial x_3} = O(\delta), \quad \frac{\partial A_2}{\partial x_3} = O(\delta), \quad \frac{\partial B_1}{\partial x_3} = O(\delta^3), \quad \frac{\partial B_2}{\partial x_3} = O(\delta^3). \tag{3–4}$$

and, by the chain rule,

$$\frac{\partial \hat{u}_1^\delta}{\partial x_2} - \frac{\partial \hat{u}_2^\delta}{\partial x_1} = \frac{\partial\,(A_1 r + B_1)}{\partial x_2} - \frac{\partial\,(A_2 r + B_2)}{\partial x_1}$$
$$= \frac{\partial (A_1 r + B_1)}{\partial \theta}\frac{\partial \theta}{\partial x_2} + \frac{\partial (A_1 r + B_1)}{\partial r}\frac{\partial r}{\partial x_2} - \frac{\partial (A_2 r + B_2)}{\partial \theta}\frac{\partial \theta}{\partial x_1} - \frac{\partial (A_2 r + B_2)}{\partial r}\frac{\partial r}{\partial x_1}$$
$$= \left(\frac{\partial A_1}{\partial \theta}\frac{\partial \theta}{\partial x_2} - \frac{\partial A_2}{\partial \theta}\frac{\partial \theta}{\partial x_1}\right) r + \left(A_1 \frac{\partial r}{\partial x_2} - A_2 \frac{\partial r}{\partial x_1}\right) + \left(\frac{\partial B_1}{\partial \theta}\frac{\partial \theta}{\partial x_2} - \frac{\partial B_2}{\partial \theta}\frac{\partial \theta}{\partial x_1}\right).$$

It follows from (3–3), (3–4) that

$$\begin{aligned}\left(\frac{\partial A_1}{\partial \theta}\frac{\partial \theta}{\partial x_2} - \frac{\partial A_2}{\partial \theta}\frac{\partial \theta}{\partial x_1}\right) r &= -\big(2a - cg'(0)\big)\sin\theta\cos\theta \\ &\quad - \big(2b - cg'(0)\big)\, r\cos\theta\left(-\frac{\sin\theta}{r}\right) + O(\delta) \\ &= 2\,(b-a)\sin\theta\cos\theta + O(\delta)\,,\end{aligned}$$

$$\begin{aligned}A_1\frac{\partial r}{\partial y} - A_2\frac{\partial r}{\partial x_1} &= (2a - cg'(0))\cos\theta\sin\theta - (2b - cg'(0))\cos\theta\sin\theta + O(\delta) \\ &= 2(a-b)\sin\theta\cos\theta + O(\delta),\end{aligned}$$

and

$$\begin{aligned}\frac{\partial B_1}{\partial \theta}\frac{\partial \theta}{\partial x_2} - \frac{\partial B_2}{\partial \theta}\frac{\partial \theta}{\partial x_1} &= \big(-2\delta^2\left(cg'(0) - a\right)\sin\theta + O\left(\delta^3\right)\big)\left(\frac{\cos\theta}{r}\right) \\ &\quad - \big(2\delta^2\left(cg'(0) - b\right)\cos\theta + O\left(\delta^3\right)\big)\left(-\frac{\sin\theta}{r}\right) \\ &= 2\delta^2\,(a-b)\sin\theta\cos\theta\frac{1}{r} + O\left(\delta^3\right)\frac{1}{r}.\end{aligned}$$

These imply that, for $\delta^2 \le r \le 2\delta^2$,

$$\frac{\partial \hat{u}_1^\delta}{\partial x_2} - \frac{\partial \hat{u}_2^\delta}{\partial x_1} = 2\delta^2(a-b)\sin\theta\cos\theta\frac{1}{r} + O(\delta).$$

Since $|\hat{u}_3^\delta|^2 = 1 - |\hat{u}_1^\delta|^2 - |u_2^\delta|^2$, we have

$$\hat{u}_3^\delta\frac{\partial \hat{u}_3^\delta}{\partial x_1} = -\hat{u}_1^\delta\frac{\partial \hat{u}_1^\delta}{\partial x_1} - \hat{u}_2^\delta\frac{\partial \hat{u}_2^\delta}{\partial x_1}, \qquad \hat{u}_3^\delta\frac{\partial \hat{u}_3^\delta}{\partial x_2} = -\hat{u}_1^\delta\frac{\partial \hat{u}_1^\delta}{\partial x_2} - \hat{u}_2\frac{\partial \hat{u}_2^\delta}{\partial x_2}$$

and $\hat{u}_3^\delta = 1 + O(\delta^2)$. Then for $\delta^2 \le r \le 2\delta^2$, we have

$$\frac{\partial \hat{u}_3^\delta}{\partial x_1} = O(\delta^2), \qquad \frac{\partial \hat{u}_3^\delta}{\partial x_2} = O(\delta^2). \tag{3–5}$$

We consider a new map $\hat{u}^\delta = \hat{u}_1^\delta I + \hat{u}_2^\delta J + \hat{u}_3^\delta K$, and thus obtain

$$\begin{aligned} &\left|\hat{u}^\delta \cdot \operatorname{curl} \hat{u}^\delta\right|^2 = 4\delta^4 (a-b)^2 \sin^2\theta \cos^2\theta \frac{1}{r^2} + O(\delta), \\ &\left|\hat{u}^\delta \times \operatorname{curl} \hat{u}^\delta\right|^2 = O(\delta). \end{aligned} \tag{3–6}$$

Moreover, we have

$$\begin{aligned} \frac{\partial \hat{u}_1^\delta}{\partial x_1} &= \frac{\partial (A_1 r + B_1)}{\partial x_1} = \left(\frac{\partial A_1}{\partial \theta} r + \frac{\partial B_1}{\partial \theta}\right)\frac{\partial \theta}{\partial x_1} + A_1 \frac{\partial r}{\partial x_1} + O(\delta) \\ &= (2a - cg'(0)) + 2\delta^2 (cg'(0) - a)\frac{\sin^2\theta}{r} + O(\delta). \end{aligned} \tag{3–7}$$

Similarly,

$$\frac{\partial \hat{u}_2^\delta}{\partial x_2} = \frac{\partial (A_2 r + B_2)}{\partial x_2} = (2b - cg'(0)) + 2\delta^2 (cg'(0) - b)\frac{\cos^2\theta}{r} + O(\delta). \tag{3–8}$$

It follows from (3–4), (3–7) and (3–8) that

$$\begin{aligned} (\operatorname{div} \hat{u}^\delta)^2 &= 4\Big(a + b - cg'(0) + \frac{\delta^2 cg'(0)}{r} - \delta^2 \frac{(a\sin^2\theta + b\cos^2\theta)}{r}\Big)^2 + O(\delta) \\ &= 4(a+b)^2 + 4c^2 (g'(0))^2 + \frac{4\delta^4 c^2 (g'(0))^2}{r^2} + \frac{4\delta^4 (a\sin^2\theta + b\cos^2\theta)^2}{r^2} \\ &\quad -8(a+b)cg'(0) + 8\delta^2 (a+b)\frac{cg'(0)}{r} - 8\delta^2 (a+b)\frac{a\sin^2\theta + b\cos^2\theta}{r} \\ &\quad - \frac{8\delta^2 c^2 (g'(0))^2}{r} + 8\delta^2 cg'(0)\frac{a\sin^2\theta + b\cos^2\theta}{r} - 8\delta^4 cg'(0)\frac{a\sin^2\theta + b\cos^2\theta}{r^2}. \end{aligned}$$

Combining this estimate with (3–6) yields

$$\begin{aligned} &\int_{\delta^2 \le r \le 2\delta^2} (\operatorname{div} \hat{u}^\delta)^2 + |\operatorname{curl} \hat{u}^\delta|^2 \, dx_1 dx_2 = \int_0^{2\pi} \int_{\delta^2}^{2\delta^2} (\operatorname{div} \hat{u}^\delta)^2 + |\operatorname{curl} \hat{u}^\delta|^2 \, r dr \, d\theta \\ &= 12\pi\delta^4 \big((a+b)^2 + c^2 (g'(0))^2 - 2(a+b)cg'(0)\big) + 16\pi\delta^4 (a+b)cg'(0) \\ &\quad -8\pi\delta^4 (a+b)^2 - 16\pi\delta^4 c^2 (g'(0))^2 + 8\pi\delta^2 cg'(0)(a+b) \\ &\quad +4\delta^4 \ln 2 \int_0^{2\pi} [c^2 (g'(0))^2 - 2cg'(0)(a\sin^2\theta + b\cos^2\theta)] \, d\theta \\ &\quad +4\delta^4 \ln 2 \int_0^{2\pi} [a^2 \sin^4\theta + b^2 \cos^4\theta + (a^2 + b^2)\sin^2\theta\cos^2\theta] \, d\theta \\ &= 4\pi\delta^4 ((a+b)^2 - g'(0)^2 c^2 + \big(a^2 + b^2 + 2(g'(0)c)^2 - 2ag'(0)c - 2bg'(0)c\big)\ln 2). \end{aligned}$$

It follows from [Brezis and Coron 1983] that

$$\int_{\delta^2\le r\le 2\delta^2} |\nabla \hat{u}^\delta|^2 dx_1dx_2 = \int\limits_{\delta^2\le r\le 2\delta^2} \big[|\hat{u}^\delta_{x_1}|^2 + |\hat{u}^\delta_{x_2}|^2 + O(\delta)\big]\, dx_1dx_2$$

$$= 4\pi\delta^4\Big(a^2+b^2-\frac{g'(0)^2c^2}{2} + \big(a^2+b^2+2(g'(0)c)^2 - 2ag'(0)c - 2bg'(0)c\big)\ln 2\Big) + O(\delta^5).$$

It follows from (3–3), (3–4) and (3–5) that

$$\begin{aligned}(3\text{–}9)\quad &\int_{\delta^2\le r\le 2\delta^2} W(\hat{u}^\delta, \nabla\hat{u}^\delta)\, dx_1dx_2\\
&= \alpha\int_{\delta^2\le r\le 2\delta^2} \left|\nabla\hat{u}^\delta\right|^2\, dx_1dx_2 + \int_{\delta^2\le r\le 2\delta^2} (k_1-\alpha)(\operatorname{div} u)^2\, dx_1dx_2\\
&\quad + \int_{\delta^2\le r\le 2\delta^2} \big[(k_2-\alpha)|\hat{u}^\delta\cdot \operatorname{curl}\hat{u}^\delta|^2 + (k_3-\alpha)|\hat{u}^\delta\times\operatorname{curl}\hat{u}^\delta|^2\big]\, dx_1dx_2\\
&= 4\alpha\pi\delta^4\Big(a^2+b^2-\frac{g'(0)^2c^2}{2} + \big(a^2+b^2+2(g'(0)c)^2-2ag'(0)c-2bg'(0)c\big)\ln 2\Big)\\
&\quad + (k_1-\alpha)4\pi\delta^4\Big((a+b)^2 - g'(0)^2c^2 + \big(a^2+b^2+2(g'(0)c)^2\big)\ln 2\Big)\\
&\quad -8\alpha\pi\delta^4\big(ag'(0)c+bg'(0)c\big)\ln 2 + \pi(k_2-k_1)\delta^4(a-b)^2\ln 2 + O(\delta^5).\end{aligned}$$

On other hand, we see

$$\frac{\partial u^\delta}{\partial x_1} = \frac{\partial\hat{u}^\delta}{\partial x_1} + O(\delta),\qquad \frac{\partial u^\delta}{\partial x_2} = \frac{\partial\hat{u}^\delta}{\partial x_2} + O(\delta)$$

and

$$\frac{\partial u^\delta}{\partial x_3} = \frac{\partial\hat{u}^\delta_1}{\partial x_3}I(x_3) + \frac{\partial\hat{u}^\delta_2}{\partial x_3}J(x_3) + \frac{\partial\hat{u}^\delta_3}{\partial x_3}K(x_3) + \hat{u}^\delta_1\frac{dI(x_3)}{dx_3} + \hat{u}^\delta_2\frac{dJ(x_3)}{dx_3} + \hat{u}^\delta_3\frac{dK(x_3)}{dx_3}.$$

For $\delta^2 \le r \le 2\delta^2$, we have

$$\frac{\partial\hat{u}^\delta_1}{\partial x_3} = O(\delta^2),\quad \frac{\partial\hat{u}^\delta_2}{\partial x_3} = O(\delta^2),\quad \frac{\partial\hat{u}^\delta_3}{\partial x_3} = O(\delta^4),$$

and, moreover,

$$\hat{u}^\delta(x_1,x_2,x_3) = \hat{u}^\delta(0,0,x_3) + O(\delta^2) = K(x_3) + O(\delta^2).$$

It follows from $u(0,0,x_3) = K(x_3)$ that for $\delta^2\le r\le 2\delta^2$, we have

$$\frac{\partial u^\delta}{\partial x_3}(x_1,x_2,x_3) = \frac{\partial u}{\partial x_3}(0,0,x_3) + O(\delta^2).$$

Using $|u| = 1$ and $u_3(0,0,0) = (0,0,1)$, we have $\partial u_3/\partial x_3(0,0,0) = 0$. Thus

$$\frac{\partial u_3^\delta}{\partial x_3}(x_1, x_2, x_3) = \frac{\partial u_3}{\partial x_3}(0,0,x_3) + O(\delta^2) = O(\delta).$$

For $\delta^2 \le r \le 2\delta^2$, we have

$$(\text{div } u^\delta)^2 = (\text{div } \hat{u}^\delta)^2 + O(\delta), \quad (u^\delta \cdot \text{curl}\, u^\delta)^2 = (\hat{u}^\delta \cdot \text{curl}\, \hat{u}^\delta)^2 + O(\delta) \tag{3–10}$$

and

$$\begin{aligned}(u^\delta \times \text{curl}\, u^\delta)^2 &= (\hat{u}^\delta \times \text{curl}\, \hat{u}^\delta)^2 + \left|\frac{\partial u_1^\delta}{\partial x_3}\right|^2 + \left|\frac{\partial u_2^\delta}{\partial x_3}\right|^2 + O(\delta) \\ &= (\hat{u}^\delta \times \text{curl}\, \hat{u}^\delta)^2 + d^2 + f^2 + O(\delta),\end{aligned} \tag{3–11}$$

where we have set

$$d = \frac{\partial u_1}{\partial x_3}(0,0,0), \quad f = \frac{\partial u_2}{\partial x_3}(0,0,0).$$

It follows from (3–9), (3–10)-(3–11) that

$$\begin{aligned}&\int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{\delta^2 \le r \le 2\delta^2} dx_1 dx_2\, W(u^\delta, \nabla u^\delta) \\ &= \int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{\delta^2 \le r \le 2\delta^2} dx_1 dx_2\, W(\hat{u}^\delta, \nabla \hat{u}^\delta) + 6\pi k_3 \delta^5 (d^2 + f^2) + O(\delta^6) \\ &= 8\alpha\pi\delta^5 \Big(a^2 + b^2 - \frac{g'(0)^2 c^2}{2} + \big(a^2 + b^2 + 2(g'(0)c)^2 - 2ag'(0)c - bg'(0)c\big) \ln 2\Big) \\ &\quad + 8\pi\delta^5(k_1 - \alpha) \Big((a+b)^2 - g'(0)^2 c^2 + \big(a^2 + b^2 + 2(g'(0)c)^2\big) \ln 2\Big) \\ &\qquad - 16\pi\delta^5(k_1 - \alpha)\big(ag'(0)c + bg'(0)c\big) \ln 2 + 2\pi(k_2 - k_1)\delta^5(a-b)^2 \ln 2 \\ &\qquad + 6\pi k_3 \delta^5 (d^2 + f^2) + O(\delta^6).\end{aligned}$$

Next, we estimate $\int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{0 \le r \le \delta^2} dx_1 dx_2\, W(u^\delta, \nabla u^\delta)$.

Let u_0 be the map defined by $u_0 = \tilde{u}(x_1, x_2)$ in the Section 2. By (2–1) and (2–2), we know

$$\begin{aligned}&\int_{r \le \delta^2} W(\hat{u}^\delta, \nabla \hat{u}^\delta)\, dx_1 dx_2 = \int_{r \le 1/c\delta^2} W(u_0, \nabla u_0)\, dx_1 dx_2 \\ &= 4\pi k_1 \int_0^{1/c\delta^2} \frac{g^2(r)}{r}\, dr + 4\pi(k_1 - \alpha) \int_0^{1/c\delta^2} gg'\, dr \\ &= 4\pi k_1 \int_0^\infty \frac{g^2(r)}{r}\, dr - 4\pi k_1 \int_{1/c\delta^2}^\infty \frac{g^2}{r^2}\, r\, dr + 2\pi(k_1 - \alpha)\delta^4 c^2 (g'(0))^2 + O(\delta^5) \\ &= 8\pi\Gamma(k_1, k_2, k_3) - 2\pi\alpha\delta^4 c^2 [g'(0)]^2 + O(\delta^5),\end{aligned} \tag{3–12}$$

because $g(r) = g'(0)r + O(r^2)$ and $g(r) = g(1/r)$.

On other hand, we see

$$\frac{\partial u^\delta}{\partial x_1} = \frac{\partial \hat{u}^\delta}{\partial x_1} + O(\delta), \quad \frac{\partial u^\delta}{\partial x_2} = \frac{\partial \hat{u}^\delta}{\partial x_2} + O(\delta) \tag{3–13}$$

and

$$\frac{\partial u^\delta}{\partial x_3} = \hat{u}_1^\delta \frac{dI(x_3)}{dx_3} + \hat{u}_2^\delta \frac{dJ(x_3)}{dx_3} + \hat{u}_3^\delta \frac{dK(x_3)}{dx_3}, \tag{3–14}$$

which implies

$$\frac{\partial u_1^\delta}{\partial x_3} = \hat{u}_2^\delta \frac{dJ(x_3)}{dx_3} \cdot I + \hat{u}_3^\delta \frac{dK(x_3)}{dx_3} \cdot I + O(\delta),$$
$$\frac{\partial u_2^\delta}{\partial x_3} = \hat{u}_1^\delta \frac{dI(x_3)}{dx_3} \cdot J + \hat{u}_3^\delta \frac{dK(x_3)}{dx_3} \cdot J + O(\delta).$$

From the results in the Section 2, we have

$$|\nabla \hat{u}^\delta| \le C \frac{g(r/\lambda)}{r} \le C \frac{\delta^4}{\delta^8 + r^2}. \tag{3–15}$$

We estimate the term

$$u_1^\delta \frac{\partial u_3^\delta}{\partial x_2} \frac{\partial \hat{u}_2^\delta}{\partial x_3} = \left(\hat{u}_1^\delta + O(\delta)\right) \left(\frac{\partial \hat{u}_3^\delta}{\partial x_2} + O(\delta)\right) \left(\hat{u}_1^\delta \frac{dI(x_3)}{dx_3} \cdot J + \hat{u}_3^\delta \frac{dK(x_3)}{dx_3} \cdot J + O(\delta)\right)$$
$$= \left(-\frac{g^2(\rho)\, g'(\rho) \cos\theta \sin\theta}{\lambda\sqrt{1 - g(\rho)^2}}\right) \left(g(\rho) \sin\theta \frac{dJ(x_3)}{dx_3} \cdot I + \sqrt{1 - g(\rho)^2} \frac{dK(x_3)}{dx_3} \cdot I\right)$$
$$+ O(\delta) \left|\frac{\partial \hat{u}_3^\delta}{\partial x_2}\right| + O(\delta),$$

where $\rho = r/\lambda$. Integrating the above identity and using (3–15), we have

$$\int_{r \le \delta^2} u_1^\delta \frac{\partial u_3^\delta}{\partial x_2} \frac{\partial u_2^\delta}{\partial x_3}\, dx_1 dx_2 = \int_0^{\delta^2} \int_0^{2\pi} u_1^\delta \frac{\partial u_3^\delta}{\partial x_2} \frac{\partial u_2^\delta}{\partial x_3}\, d\theta\, r\, dr = O(\delta^5 \ln(1/\delta)).$$

Similarly, we obtain

$$\int_{r \le \delta^2} u_2^\delta \frac{\partial u_3^\delta}{\partial x_1} \frac{\partial u_1^\delta}{\partial x_3}\, dx_1 dx_2 = O(\delta^5 \ln \delta).$$

By a similar argument, we also have

$$\int_{r \le \delta^2} \frac{\partial u_3^\delta}{\partial x_1} \frac{\partial u_1^\delta}{\partial x_3}\, dx_1 dx_2 = O(\delta^5 \ln \delta), \qquad \int_{r \le \delta^2} \frac{\partial u_3^\delta}{\partial x_2} \frac{\partial u_2^\delta}{\partial x_3}\, dx_1 dx_2 = O(\delta^5 \ln \delta).$$

From (3–14), it is easy to see that

$$\int_{r\le\delta^3}\left|\frac{\partial u^\delta}{\partial x_3}(x_1,x_2,x_3)\right|^2 dx_1dx_2 = O(\delta^6).$$

By combining above estimates with the identity $\left|\operatorname{curl} u^\delta\right|^2 = \left|u\cdot\operatorname{curl} u^\delta\right|^2 + \left|u\times\operatorname{curl} u^\delta\right|^2$, we obtain

$$\int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{r\le\delta^3} dx_1dx_2\, W(u^\delta,\nabla u^\delta) = \int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{r\le\delta^3} dx_1dx_2\, W(\hat u^\delta,\nabla \hat u^\delta) + O(\delta^6\ln\delta). \tag{3–16}$$

For $\delta^3\le r\le\delta^2$, it follows from (2–5) and (2–7) that

$$g(r/\lambda)\le C\frac{r\lambda}{\lambda^2+r^2} = O(\delta) \tag{3–17}$$

for some constant C. Using (3–17), it follows from (3–13)-(3–14) that

$$\int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{\delta^3\le r\le\delta^2} dx_1dx_2\, W(u^\delta,\nabla u^\delta)$$
$$=\int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{\delta^3\le r\le\delta^2} dx_1dx_2 \left(W(\hat u^\delta,\nabla \hat u^\delta)+k_3\left[\left|\frac{\partial u_1^\delta}{\partial x_3}\right|^2+\left|\frac{\partial u_2^\delta}{\partial x_3}\right|^2\right]\right)+O(\delta^6\ln\delta).$$

Combining this with (3–12)and (3–16) yields

$$\int_{-\delta+\delta^2}^{\delta-\delta^2} dx_3 \int_{r\le\delta^2} dx_1dx_2\, W(u^\delta,\nabla u^\delta) = 16\pi(\delta-\delta^2)\Gamma(k_1,k_2,k_3) - 4\pi\alpha\delta^5c^2[g'(0)]^2$$
$$+2\pi k_3\delta^5(d^2+f^2)+O(\delta^6\ln\delta).$$

Case 2. Estimate for $E(u^\delta)$ in c_P^δ and c_N^δ.

Let G_P be the little cone inside c_P^δ with vertex $P=(0,0,\delta)$ given by

$$G_P=\left\{(x_1,x_2,x_3)\in\mathbb{R}^3 \;\middle|\; (x_1)^2+(x_2)^2\le(\delta-x_3)^2,\quad \delta-\delta^2\le x_3\le\delta\right\}.$$

Its end is the disk

$$D_{\delta^2}=\left\{(x_1',x_2',x_3')\in\mathbb{R}^3 \;\middle|\; r'^2=x'^2+y'^2\le\delta^4,\quad x_3'=\delta-\delta^2\right\}.$$

Let $x=(x_1,x_2,x_3)$ be a point in G_P and let $x'=(x_1',x_2',x_3')$ be its projection $x'=\Pi^+(x)$ on the disk D_{δ^2} as

$$x'=\Pi^+(x)=\left(\frac{\delta^2x_1}{\delta-x_3},\ \frac{\delta^2x_2}{\delta-x_3},\ \delta-\delta^2\right).$$

Now u^δ is constant on the rays passing by P, that is,

$$u^\delta(x_1, x_2, x_3) = u^\delta(x_1', x_2', \delta - \delta^2).$$

Using the chain rule, it follows from the previous two equations that for a point $(x_1, x_2, x_3) \in G_P$,

$$\begin{aligned}
&\frac{\partial u^\delta}{\partial x_1}(x_1, x_2, x_3) = \frac{\delta^2}{\delta - x_3}\,\frac{\partial u^\delta}{\partial x_1'}(x_1', x_2', \delta - \delta^2)\\
&\frac{\partial u^\delta}{\partial x_2}(x_1, x_2, x_3) = \frac{\delta^2}{\delta - x_3}\,\frac{\partial u^\delta}{\partial x_2'}(x_1', x_2', \delta - \delta^2)\\
&\frac{\partial u^\delta}{\partial x_3}(x_1, x_2, x_3) = \frac{x_1'}{\delta - x_3}\,\frac{\partial u^\delta}{\partial x_1'}(x_1', x_2', \delta - \delta^2) + \frac{x_2'}{\delta - x_3}\,\frac{\partial u^\delta}{\partial x_2'}(x_1', x_2', \delta - \delta^2).
\end{aligned} \tag{3–18}$$

Using the third identity of (3–18), we have

$$\begin{aligned}
&\int_{G_P} \Big|\frac{\partial u^\delta}{\partial x_3}(x_1, x_2, x_3)\Big|^2 dx_1dx_2dx_3\\
&= \int_{\delta-\delta^2}^{\delta} dx_3 \int_{r^2 \le (\delta - x_3)^2} dx_1dx_2\, \frac{1}{(\delta - x_3)^2}\left(x_1'\frac{\partial u^\delta}{\partial x_1'} + x_2'\frac{\partial u^\delta}{\partial x_2'}\right)^2 (x_1', x_2', \delta - \delta^2)\\
&= \int_{r' \le \delta^2} \frac{1}{\delta^2}\left(x_1'\frac{\partial u^\delta}{\partial x_1'} + x_2'\frac{\partial u^\delta}{\partial x_2'}\right)^2 (x_1', x_2', \delta - \delta^2)\, dx_1'dx_2'.
\end{aligned}$$

On other hand, from the results in the Section 2, we obtain

$$\begin{aligned}
&|\nabla_{x_1'} u^\delta|^2(x_1', x_2', \delta - \delta^2) + |\nabla_{x_2'} u^\delta|^2(x_1', x_2', \delta - \delta^2)\\
&\qquad\qquad \le C\frac{g(r'/c\delta^2)}{r'^2} \le C\frac{\lambda^2}{(\lambda^2 + r'^2)^2} \le C\frac{\delta^8}{(\delta^8 + r'^2)^2},
\end{aligned}$$

where $\lambda = c\delta^4$. Combining the previous two equations, we obtain

$$\int_{G_P} \Big|\frac{\partial u^\delta}{\partial x_3}(x_1, x_2, x_3)\Big|^2 dx_1dx_2dx_3 \le C\int_0^{\delta^2} \frac{\delta^6 r'^2}{(\delta^8 + r'^3)^2}\, dr' = O(\delta^6 \ln(\delta)).$$

A simple calculation yields

$$\begin{aligned}
&\left(u_2^\delta\, \frac{\partial u_3^\delta}{\partial x_1'}\Big(x_1'\, \frac{\partial u_1^\delta}{\partial x_1'} + x_2'\frac{\partial u_1^\delta}{\partial x_2'}\Big)\right)(x_1', x_2', \delta - \delta^2)\\
&\quad = \left(g^2(r'/\lambda) + g(r'/\lambda)\Big(\frac{1}{r'}g(r'/\lambda)\Big)_{r'}\right)\left(\pm\sqrt{1 - g^2(r'/\lambda)}\right)_{r'} \sin\theta\cos\theta,
\end{aligned} \tag{3–19}$$

where we use polar coordinates $x_1' = r'\cos\theta$ and $x_2' = r'\sin\theta$.

By (3–18) and (3–19), we calculate

$$\int_{\delta-\delta^2}^{\delta} dx_3 \int_{r\le\delta-x_3} dx_1dx_2\, u_2^\delta(x_1,x_2,x_3)\frac{\partial u_3^\delta}{\partial x_1}(x_1,x_2,x_3)\frac{\partial u_1^\delta}{\partial x_3}(x_1,x_2,x_3)$$
$$=\frac{1}{\delta^2}\int_{\delta-\delta^2}^{\delta} dx_3 \int_{r'\le\delta^2} dx_1dx_2\,\Big(u_2^\delta\frac{\partial u_3^\delta}{\partial x_1'}\big(x_1'\frac{\partial u_1^\delta}{\partial x_1'}+x_2'\frac{\partial u_1^\delta}{\partial x_2'}\big)\Big)(x_1',x_2',\delta-\delta^2)=0.$$

Similarly, we have

$$\int_{\delta-\delta^2}^{\delta} dx_3 \int_{r\le\delta-x_3} dx_1dx_2\,\frac{\partial u_3^\delta}{\partial x_1}(x_1,x_2,x_3)\frac{\partial u_1^\delta}{\partial x_3}(x_1,x_2,x_3)=0,$$
$$\int_{\delta-\delta^2}^{\delta} dx_3 \int_{r\le\delta-x_3} dx_1dx_2\, u_1^\delta(x_1,x_2,x_3)\frac{\partial u_3^\delta}{\partial x_2}(x_1,x_2,x_3)\frac{\partial u_2^\delta}{\partial x_3}(x_1,x_2,x_3)=0,$$
$$\int_{\delta-\delta^2}^{\delta} dx_3 \int_{r\le\delta-x_3} dx_1dx_2\,\frac{\partial u_3^\delta}{\partial x_2}(x_1,x_2,x_3)\frac{\partial u_2^\delta}{\partial x_3}(x_1,x_2,x_3)=0.$$

Combining these estimates with (3–12) yields

$$\int_{G_P} W(u^\delta,\nabla u^\delta)(x_1,x_2,x_3)\,dx_1dx_2dx_3$$
$$=\int_{\delta-\delta^2}^{\delta} dx_3' \int_{r'\le\delta^2} dx_1'\,dx_2\,W(u^\delta,\nabla u^\delta)(x_1',x_2',\delta-\delta^2)+O(\delta^6\ln\delta)$$
$$=8\pi\Gamma(k_1,k_2,k_3)\delta^2+O(\delta^6\ln\delta).$$

Since u is regular and $|\nabla u^\delta|$ is bounded by a constant, we obtain

$$\int_{c_P^\delta\setminus G_P} W(u^\delta,\nabla u^\delta)\,dx_1dx_2dx_3=O(\delta^6).$$

Therefore, it follows from the previous two equations that

$$\int_{c_P^\delta} W(u^\delta,\nabla u^\delta)\,dx_1dx_2dx_3=8\pi\delta^2\Gamma(k_1,k_2,k_3)+O(\delta^6\ln\delta).$$

Similarly, we get $\int_{c_N^\delta} W(u^\delta,\nabla u^\delta)\,dx_1dx_2dx_3=8\pi\delta^2\Gamma(k_1,k_2,k_3)+O(\delta^6\ln\delta)$.

Proof of Theorem 1.2 for $k_2\ge k_1$. Since u is smooth, we have

$$\int_{-\delta-\delta^2}^{\delta+\delta^2} dx_3 \int_{r\le2\delta^2} dx_1dx_2\,W(u,\nabla u)(x_1,x_2,x_3)$$
$$=\int_{-\delta-\delta^2}^{\delta+\delta^2} dx_3 \int_{r\le2\delta^2} dx_1dx_2\,\big[\alpha(a^2+b^2)+(k_1-\alpha)(a+b)^2+k_3(d^2+f^2)\big]+O(\delta^6)$$
$$=8\pi\delta^5\big[\alpha(a^2+b^2)+(k_1-\alpha)(a+b)^2+k_3(d^2+f^2)\big]+O(\delta^6).$$

Also

$$\int_{c_N^\delta} W(u,\nabla u)\,dx_1dx_2dx_3 = O(\delta^6), \qquad \int_{c_P^\delta} W(u,\nabla u)\,dx_1dx_2dx_3 = O(\delta^6).$$

Finally, we have

$$\begin{aligned} &E\left(u^\delta,\Omega\right) - E(u,\Omega) - 16\pi\Gamma(k_1,k_2,k_3)\delta \\ &\quad = -8k_1\pi\delta^5\Big(g'(0)^2c^2 - \big(a^2+b^2+2(g'(0)c)^2 - 2ag'(0)c - bg'(0)c\big)\ln 2\Big) \\ &\qquad\qquad + 2\pi\delta^5(k_2-k_1)(a-b)^2\ln 2 + O(\delta^6\ln\delta). \end{aligned}$$

When $k_2 - k_1 \le k_1 4(1-\ln 2)/\ln 2$, we choose $g'(0)c = \max\{a,b\}$ to obtain

$$g'(0)^2c^2 - \big(a^2+b^2+2(g'(0)c)^2 - (2a+b)g'(0)c\big)\ln 2 - \frac{k_2-k_1}{4k_1}(a-b)^2\ln 2 > 0.$$

Choosing δ sufficiently small, Theorem 1.2 is proved. □

Remark 3.1. Let $u_0 \in H^1_\gamma(\Omega, S^2)$ for smooth boundary data γ with $\deg(\gamma) \neq 0$. For a map $u \in H^1_\gamma(\Omega, S^2)$, in similar fashion to arguments of [Giaquinta et al. 1998, Chapter 4], there is a one-dimensional rectifiable current L_{u,u_0} with $-\partial L_{u,u_0} = P(u) - P(u_0)$ that minimizes the mass among all one-dimensional rectifiable currents L with $-\partial L = P(u) - P(u_0)$, where $P(u)$ is the zero-dimensional current in Ω determined by u (see [Giaquinta et al. 1998, Chapter 4]). Moreover, we have

$$M(L_{u,u_0}) = L(u,u_0).$$

For $u \in H^1_\gamma(\Omega, S^2)$, consider

$$[\![u]\!] := \left\{ T = G_{u_T} + L_{u,u_0} \times [\![S^2]\!] \;\middle|\; T - [\![G_{u_0}]\!] \in \mathrm{Cart}^{2,1}(\Omega\times\mathbb{R}^3), \quad u_T = u \right\}.$$

Define

$$\mathscr{E}([\![u]\!]) = \int_\Omega W(u(x),\nabla u(x))\,dx + 8\pi\Gamma(k_1,k_2,k_2)M(L_{u,u_0}) = F_{u_0}(u,\Omega).$$

Then the semicontinuity $\mathscr{E}$ implies that $F_{u_0}(u,\Omega)$ is also lower semicontinuous with respect to the weak convergence in $H^1_{u_0}(\Omega, S^2)$ (see [Giaquinta et al. 1989; 1998]).

Proof of Theorem 1.1 for $k_2 \ge k_1$. If there are infinitely many distinct minimizers for E in $H^1_\gamma(\Omega, S^2)$, the proof of Theorem 1.1 is completed. Now we assume that there are only a finite number of minimizers $w_1,\dots,w_m$ for E in $H^1_\gamma(\Omega, S^2)$.

By the partial regularity of [Hardt et al. 1986], with the fact that γ is not a constant, there is a new subdomain Ω_1 of Ω such that w_1 is smooth in Ω_1, and there is some $x_0 \in \Omega_1$ with $\nabla w_1(x_0) \neq 0$. For sufficiently small ρ, it follows from taking $v_1 = w^\delta$ in Theorem 1.2 that

$$E(v_1) < F_{v_1}(w_1) = E(w_1) + 8\pi\Gamma(k_1, k_2, k_3)L(v_1, w_1).$$

Let u_1 be a minimizer of F_{v_1} in $H^1_\gamma(\Omega, S^2)$, and let u_1 be a weak solution of Equation (1–3) with boundary value γ.

For δ sufficiently small, we shall prove that u_1 is different from all minimizers w_i of E in $H^1_\gamma(\Omega, S^2)$. We have two cases.

(i) $L(w_k, w_1) = 0$ for some k. It is easy to see $L(v_1, w_1) = L(v_1, w_k)$. Noticing $E(w_1) = E(w_k)$, it follows the minimality of u_1 that

$$F_{v_1}(u_1) \leq E(v_1) < E(w_1) + 8\pi\Gamma(k_1, k_2, k_3)L(v_1, w_1) = F_{v_1}(w_k).$$

This implies that $u_1 \neq w_k$.

(ii) $L(w_k, w_1) > 0$. We know

$$L(w_k, v_1) + L(v_1, w_1) \geq L(w_k, w_1).$$

This implies

$$\begin{aligned} F_{v_1}(w_k) &= E(w_k) + 8\pi\Gamma(k_1, k_2, k_3)L(w_k, v_1) \\ &\geq E(w_1) + 8\pi\Gamma(k_1, k_2, k_3)(L(w_k, w_1) - 2\rho). \end{aligned}$$

Choose $\rho > 0$ sufficiently small so that

$$0 < 2\rho < \frac{L(w_k, w_1)}{2}.$$

This gives $L(w_k, w_1) - 2\rho > 2\rho > L(v_1, w_1)$. Therefore

$$F_{v_1}(w_k) > F_{v_1}(w_1) \geq F_{v_1}(u_1).$$

So u_1 is different from all minimizers w_k of E.

We construct by induction a sequence u_j of distinct weak solutions of (1–3) in $H^1_\gamma(\Omega, S^2)$ which are also different from the minimizer w_i. Choose ρ_{j+1} such that

$$0 < 2\rho_{j+1} < \min\{L(w_k, w_1)/2, \text{ with } L(w_k, w_1) > 0\}$$

and

$$0 < 2\rho_{j+1} < \min\left\{\frac{E(u_i) - E(w_1)}{8\pi\Gamma(k_1, k_2, k_3)}, \quad i = 1, \ldots, j\right\}. \tag{3–20}$$

By taking $\rho = \rho_{j+1}$ and $u = w_1$ in Theorem 1.2, there exists a v_{j+1} and $\delta_{j+1} \leq \rho_{j+1}$ such that

$$E(v_{j+1}) < E(w_1) + 16\pi\Gamma(k_1, k_2, k_3)\delta_{j+1}.$$

Let u_{j+1} be a minimizer of $F_{v_{j+1}}$ in $H^1_\gamma(\Omega, S^2)$. The same argument as above assures that u_{j+1} is different from all w_i. Next we prove that $u_{j+1} \neq u_i$ for all $i \leq j$. From the above estimates, we know

$$F_{v_{j+1}}(u_{j+1}) \leq F_{v_{j+1}}(v_{j+1}) = E(v_{j+1}) < E(w_1) + 16\pi\Gamma(k_1, k_2, k_3)\rho_{j+1}. \tag{3–21}$$

From (3–20) we have

$$16\pi\rho_{j+1}\Gamma(k_1, k_2, k_3) < E(u_i) - E(w_1).$$

Combining this with (3–21) yields

$$E(v_{j+1}) \leq F_{v_{j+1}}(u_{j+1}) < E(u_i),$$

which implies $u_{j+1} \neq u_i$ for $i = 2, \ldots, j$. Letting $j \to \infty$, we see that there exist infinitely many solutions $\{u_j\}_{j=1}^\infty$ of (1–3) in $H^1_\gamma(\Omega, S^2)$. This proves Theorem 1.1 for $k_2 \geq k_1$. □

4. Proof of Theorems 1.1 and 1.2 for $k_1 > k_2$

As in Section 3.1, for a sufficiently small δ and x_3 in $[-\delta + \delta^2, \delta - \delta^2]$, we may choose

$$K(x_3) = u(0, 0, x_3), \qquad I(x_3) = \frac{u_{x_1}(0, 0, x_3)}{|u_{x_1}(0, 0, x_3)|}.$$

to form a basis $\{I(x_3), J(x_3), K(x_3)\}$ of $\mathbb{R}^3$ depending on x_3. We write

$$u = \hat{u}_1 I(x_3) + \hat{u}_2 J(x_3) + \hat{u}_3 K(x_3)$$

with $\hat{u}_1(0, 0, x_3) = \hat{u}_2(0, 0, x_3) = 0$, $\hat{u}_3(0, 0, x_3) = 1$. There are two numbers $a > 0$ and $b \leq 0$ (for a suitable rotation of $\mathbb{R}^3$) such that

$$u_{x_1}(0, 0, x_3) = (a + O(x_3))J(x_3), \quad u_{x_2}(0, 0, x_3) = (b + O(x_3))I(x_3) + O(x_3)J(x_3).$$

We consider the cylinder C^δ in $\mathbb{R}^3$ defined by

$$C^\delta = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \;\middle|\; 0 \leq r \leq \delta + \delta^2, \quad -\delta - \delta^2 \leq x_3 \leq \delta + \delta^2\right\}.$$

As in Section 3.1, we construct a map $u^\delta = \hat{u}_1^\delta I(x_3) + \hat{u}_2^\delta J(x_3) + \hat{u}_3^\delta K(x_3)$ as follows:

(i) $u^\delta = u$ outside C^δ.

(ii) Inside C^δ, for each $x_3 \in [-\delta + \delta^2, \delta - \delta^2]$ (that is, the subcylinder of C^δ), we construct u^δ in three different cases:

   (a) If $r > 2\delta^2$, we set $u^\delta(x) = u(x)$.

(b) If $r < \delta^2$, we set, with $\rho = r/\lambda$,

$$u^\delta(x_1, x_2, x_3) = g(\rho)\frac{x_2}{r} I(x_3) - g(\rho)\frac{x_1}{r} J(x_3) + \operatorname{sign}(1-r)\sqrt{1-g^2(\rho)}\, K(x_3).$$

(c) If $\delta^2 \le r \le 2\delta^2$, we set

$$u^\delta(x) = (A_1 r + B_1)\, I(x_3) + (A_2 r + B_2)\, J(x_3) + \sqrt{1-(A_1 r+B_1)^2-(A_2 r+B_2)^2}\, K(x_3),$$

where A_1, A_2, B_1, B_2 depend only on θ, δ, and x_3 and are determined by

$$\begin{aligned} 2\delta^2 A_1 + B_1 &= \hat{u}_1(2\delta^2\cos\theta, 2\delta^2\sin\theta, x_3),\\ 2\delta^2 A_2 + B_2 &= \hat{u}_2(2\delta^2\cos\theta, 2\delta^2\sin\theta, x_3),\\ \delta^2 A_1 + B_1 &= g(\delta^2/\lambda)\sin\theta = g(\lambda/\delta^2)\sin\theta,\\ \delta^2 A_2 + B_2 &= -g(\delta^2/\lambda)\cos\theta = -g(\lambda/\delta^2)\cos\theta, \end{aligned} \tag{4–1}$$

and $g(r)$ is the solution of Equation (2–8) with $g(0) = 0$, $g'(0) > 0$, $g(1) = 1$, $g(r) = g(1/r)$ and $\lambda = c\delta^4$, where c will be determined later.

(iii) Inside C^δ, for each

$$x_3 \in [-\delta, -\delta+\delta^2] \cup [\delta, \delta-\delta^2],$$

we let $P = (0, 0, \delta)$ and $N = (0, 0, -\delta)$, in a small cylinder c_P^δ (or c_N^δ). The cylinder is centered at P (or N) with radius $2\delta^2$, length $2\delta^2$, and its axis along the x_3-axis. If we denote by Π^+ (or Π_-) the radial projection centered at P (or N) onto the boundary of c_P^δ (or c_N^δ), the transformed map u^δ is the composition of Π^+ (or Π_-) and the value of u^δ on this boundary.

The proof of Theorem 1.2 for $k_2 \le k_1$ is very similar to the one for $k_1 \le k_2$ on page 195. We only need to make a few modifications.

For $\delta^2 \le r \le 2\delta^2$ and for each

$$x_3 \in [-\delta+\delta^2, \delta-\delta^2],$$

we solve Equation (4–1) for A_1, A_2, B_1 and B_2 to obtain

$$\begin{aligned} A_1 &= (2a - cg'(0))\sin\theta + O(\delta),\\ A_2 &= (2b + cg'(0))\cos\theta + O(\delta),\\ B_1 &= 2\delta^2(g'(0)c - a)\sin\theta + O(\delta^3),\\ B_2 &= -2\delta^2(g'(0)c + b)\cos\theta + O(\delta^3). \end{aligned} \tag{4–2}$$

and

$$
\begin{aligned}
\frac{\partial A_1}{\partial \theta} &= (2a - cg'(0))\cos\theta + O(\delta),\\
\frac{\partial A_2}{\partial \theta} &= -(2b + cg'(0))\sin\theta + O(\delta),\\
\frac{\partial B_1}{\partial \theta} &= 2\delta^2(cg'(0) - a)\cos\theta + O(\delta^3),\\
\frac{\partial B_2}{\partial \theta} &= 2\delta^2(cg'(0) + b)\sin\theta + O(\delta^3).
\end{aligned}
\tag{4–3}
$$

Using (4–2) and (4–3), we have

$$
\begin{aligned}
\frac{\partial \hat{u}_1^\delta}{\partial x_1} &= \Big(\frac{\partial A_1}{\partial \theta} r + \frac{\partial B_1}{\partial \theta}\Big)\frac{\partial \theta}{\partial x_1} + A_1\frac{\partial r}{\partial x_1} + O(\delta)\\
&= -2\delta^2(cg'(0) - a)\frac{\cos\theta\sin\theta}{r} + O(\delta),\\
\frac{\partial \hat{u}_2^\delta}{\partial x_2} &= \frac{\partial (A_2 r + B_2)}{\partial x_2} = 2\delta^2(cg'(0) + b)\frac{\sin\theta\cos\theta}{r} + O(\delta),\\
\frac{\partial \hat{u}_1^\delta}{\partial x_2} &= \frac{\partial (A_1 r + B_1)}{\partial x_2} = (2a - cg'(0)) + 2\delta^2(cg'(0) - a)\frac{\cos^2\theta}{r} + O(\delta),\\
\frac{\partial \hat{u}_2^\delta}{\partial x_1} &= \frac{\partial (A_2 r + B_2)}{\partial x_1} = (2b + cg'(0)) - 2\delta^2(cg'(0) + b)\frac{\sin^2\theta}{r} + O(\delta).
\end{aligned}
$$

Consider a new map $\hat{u}^\delta = \hat{u}_1^\delta I + \hat{u}_2^\delta J + \hat{u}_3^\delta K$. Then we have

$$
\begin{aligned}
&\big|\operatorname{div}\hat{u}^\delta\big|^2 = 4\delta^4(a+b)^2\sin^2\theta\cos^2\theta\frac{1}{r^2} + O(\delta),\\
&\big|\hat{u}^\delta \times \operatorname{curl}\hat{u}^\delta\big|^2 = O(\delta),\\
&(\operatorname{curl}\hat{u}^\delta)^2 = \Big[2(a-b) - 2cg'(0) + \frac{2\delta^2 cg'(0)}{r} - 2\delta^2\frac{(a\cos^2\theta - b\sin^2\theta)}{r}\Big]^2 + O(\delta).
\end{aligned}
$$

Using this, we have

$$
\begin{aligned}
&\int_{\delta^2\le r\le 2\delta^2} (\operatorname{div}\hat{u}^\delta)^2 + |\operatorname{curl}\hat{u}^\delta|^2\, dx_1 dx_2\\
&\quad = 4\pi\delta^4\Big((a-b)^2 - g'(0)^2c^2 + \big(a^2 + b^2 + 2(g'(0)c)^2 - (2a-2b)g'(0)c\big)\ln 2\Big).
\end{aligned}
$$

We know that

$$
\begin{aligned}
|\hat{u}^\delta_{x_1}|^2+|\hat{u}^\delta_{x_2}|^2 = 4\delta^4(g'(0)c-a)^2\frac{\cos^2\theta}{r^2} + 4\delta^4(g'(0)c+b)^2\frac{\sin^2\theta}{r^2} + (2a-g'(0)c)^2 \\
+(2b+g'(0)c)^2 + 4\delta^2(2a-g'(0)c)(g'(0)c-a)\frac{\cos^2\theta}{r} \\
-4\delta^2(2b+g'(0)c)(g'(0)c+b)\frac{\sin^2\theta}{r}.
\end{aligned}
$$

Using this, we have

$$
\begin{aligned}
&\int_{\delta^2\le r\le 2\delta^2} |\nabla\hat{u}^\delta|^2 dx_1dx_2 = \int_{\delta^2\le r\le 2\delta^2} \left[|\hat{u}^\delta_{x_1}|^2+|\hat{u}^\delta_{x_2}|^2+O(\delta)\right] dx_1dx_2 \\
&\quad = 4\pi\delta^4\Big(a^2+b^2-\frac{g'(0)^2c^2}{2} + \big(a^2+b^2+2(g'(0)c)^2-2(a-b)g'(0)c\big)\ln 2\Big) \\
&\qquad +O(\delta^5).
\end{aligned}
$$

From arguments similar to those in Section 3.2, we finally have

$$
\begin{aligned}
&E(u^\delta,\Omega)-E(u,\Omega)-16\pi\Gamma(k_1,k_2,k_3)\delta \\
&\quad = -8k_1\pi\delta^5\Big(g'(0)^2c^2-\big(a^2+b^2+2(g'(0)c)^2-2(a-b)g'(0)c\big)\ln 2\Big) \\
&\qquad +2\pi\delta^5(k_2-k_1)(a+b)^2\ln 2+O(\delta^6\ln\delta).
\end{aligned}
$$

When $0\le k_1-k_2\le k_2 4(1-\ln 2)/\ln 2$, we choose $g'(0)c=\max\{a,-b\}$ to obtain

$$
g'(0)^2c^2-\big(a^2+b^2+2(g'(0)c)^2-2(a-b)g'(0)c\big)\ln 2-\frac{k_1-k_2}{4k_1}(a+b)^2\ln 2>0.
$$

Theorem 1.2 follows from choosing δ sufficiently small.

The proof of Theorem 1.1 for the case $k_1>k_2$ is the same as one for the case $k_2\ge k_1$ in Section 3.2. We omit details here. □

5. Partial regularity of the weak solutions

We will now complete a proof of Theorem 1.3. We recall that

$$
W(u,p)=\alpha|p|^2+(k_1-\alpha)(\operatorname{tr} p)^2+(k_2-\alpha)(g\cdot u)^2+(k_3-\alpha)|g\times u|^2,
$$

where $p=(p_i^j)_{3\times 3}$ and g is the axial vector of $p-p^T$, that is, the vector defined in coordinates by

$$
g_i=\varepsilon_{ijk}p_j^k
$$

with ε_{ijk} being the components of the Levi-Civita tensor. For simplicity, we assume $\alpha=1$.

There exists a positive constant $\Lambda>0$ such that

$$
|p|^2\le W(u,p)\le\Lambda|p|^2.
$$

Lemma 5.1. *For any $\lambda \in [0, 1)$, let u_λ be a minimizer of E_λ in $H^1_\gamma(\Omega, S^2)$. Then u_λ is a quasiminimizer of the functional E in $H^1_\gamma(\Omega, S^2)$, that is, $E(u_\lambda, B) \le QE(w; B)$ for any $w \in H^1_{u_\lambda}(B, S^2)$ and any subdomain $B \subset \Omega$ with $Q = (1+\lambda)/(1-\lambda)$.*

Proof. Let R^∞_γ be a set of all maps in $H^1_\gamma(\Omega, S^2)$ having a finite number of singular points, of which $\{P_i\}$ are of positive degree $+1$ and $\{N_i\}$ are of negative degree -1 inside Ω. Let u be a map in R^∞_g.

As in [Giaquinta et al. 1989; 1998], the function

$$\Gamma(n, \xi) := \inf\bigl\{W(n, G) \bigm| M_2(G) = \xi, \quad G^T n = 0\bigr\}$$

is given at every $n \in S^2$ and $\xi = t \wedge \epsilon(n) \in \wedge_3(\mathbb{R}^3 \times T_n S^2)$, where $|t| = 1$, $\epsilon(n)$ is the unit 2-vector associated to $T_n S^2$, and G^T is the transpose of the matrix G. A calculation (see [Giaquinta et al. 1989; 1998]) yields

$$\Gamma(n, \xi) = 2\sqrt{k^2(t, n)^2 + kk_3(1-(t, n)^2)^2} + (k-\alpha)(t, n).$$

Thus

$$W(u(x), \nabla u(x)) \ge \Gamma(n, M_2(\nabla u(x)) = \Gamma\left((n, \frac{D(u(x))}{|D(u(x))|} \wedge \epsilon(n)\right) |M_2(\nabla u(x)|.$$

Integrating over B and using the co-area formula, we then have

$$\begin{aligned} E(u, B) &\ge \int_B \Gamma\left(n, \frac{D(u(x))}{|D(u(x))|} \wedge \epsilon(n)\right) |M_2(\nabla u(x)| \, dx \\ &\ge \int_{S^2} d\mathcal{H}^2(n) \int_{u^{-1}(n)} \Gamma\left(n, \frac{D(u(x))}{|D(u(x))|} \wedge \epsilon(n)\right) d\mathcal{H}^1. \end{aligned}$$

We know that $u^{-1}(n)$ is the union of curves of two kind of curves oriented by $D(u)/|D(u)|$:

(i) closed curves $\Gamma^u_1 \cup \Gamma^u_2 \cup \cdots \cup \Gamma^u_l$;

(ii) curves joining $\partial B \cap \{P^u_i\}^k_{i=1} \cap \{N^u_i\}^k_{i=1}$, where $\{P^u_i, N^u_i\}^k_{i=1}$ are all singularities of u inside B.

For any positive singularity P^u_i, there is a curve $C_i(u)$ joining P^u_i to another point $\tilde{N}^u_i$, which is either a negative singularity of the map u or a point y_i on the boundary ∂B.

Since $\Gamma(n, \cdot \wedge \epsilon(n))$ is convex and one-homogeneous, Jensen's inequality implies

$$\int_{C_i} \Gamma\left(n, \frac{D(u(x))}{|D(u(x))|} \wedge \epsilon(n)\right) d\mathcal{H}^1 \ge \Gamma\left(n, \int_{C_i} \frac{D(u(x))}{|D(u(x))|} d\mathcal{H}^1 \wedge \epsilon(n)\right)$$

Note that for any vector t with $|t| = 1$, we have [Giaquinta et al. 1990]

$$\int_{S^2} \Gamma(n, t \wedge \epsilon(n))\, d\mathcal{H}^2(n) = 8\pi\Gamma(k_1, k_2, k_3).$$

Let $w \in R^\infty_\gamma$ and $w - u \in H^1_0(B, \mathbb{R}^3)$. Then $w^{-1}(n)$ is the union of curves of two kind of curves oriented by $D(w)/|D(w)|$:

(iii) closed curves $\Gamma^w_1 \cup \Gamma^w_2 \cup \cdots \cup \Gamma^w_l$;

(iv) curves joining $\partial B \cap \{P^w_i\}^k_{i=1} \cap \{N^w_i\}^m_{i=1}$, where $\{P^w_i, N^w_i\}^m_{i=1}$ are all singularities of u inside B.

If $\tilde{N}^u_i$ is a boundary point y_i with $u(y_i) = n$ joining a curve to a positive singularity P^u_i by a curve $C^i(u)$ inside the set $u^{-1}(n)$, there is a positive singularity P^w_i of w joining to y_i by a curve $C^i(w)$ inside the set $w^{-1}(n)$. As in [Giaquinta et al. 1998], we note that $D(u(x))$ is the tangent to the level line $u(x) = n$. For an oriented curve $C_i(u)$ joining P^u_i to $\tilde{N}^u_i$ and a curve $C_i(w)$ joining P^w_i to $\tilde{N}^w_i$, we have

$$\int_{C_i(u)} \frac{D(u(x))}{|D(u(x))|}\, d\mathcal{H}^1 = -(P^u_i - \tilde{N}^u_i), \quad \int_{C_i(w)} \frac{D(w(x))}{|D(w(x))|}\, d\mathcal{H}^1 = -(P^w_i - \tilde{N}^w_i),$$

where $\tilde{N}^u_i$ is either a negative singularity of u or a boundary y_i with $u(y) = n$, and $\tilde{N}^w_i$ is either a negative singularity of w or a boundary y_i with $w(y) = n$.

Then

$$\begin{aligned}
E(u, B) + E(w, B) &\geq \int_{S^2} d\mathcal{H}^2(n) \sum_{i=1}^{k} \Gamma\left(n, \int_{C_i(u)} \frac{D(u(x))}{|D(u(x))|} d\mathcal{H}^1 \wedge \epsilon(n)\right) \\
&\quad + \int_{S^2} d\mathcal{H}^2(n) \sum_{i=1}^{m} \Gamma\left(n, \int_{C_i(w)} \frac{D(w(x))}{|D(w(x))|} d\mathcal{H}^1 \wedge \epsilon(n)\right) \\
&= \sum_{i=1}^{k} |P^u_i - \tilde{N}^u_i| \int_{S^2} \Gamma\left(n, \frac{P^u_i - \tilde{N}^u_i}{|p^u_i - \tilde{N}^u_i|} \wedge \epsilon(n)\right) d\mathcal{H}^2(n) \\
&\quad + \sum_{i=1}^{k} |P^w_i - \tilde{N}^w_i| \int_{S^2} \Gamma\left(n, \frac{P^w_i - \tilde{N}^w_i}{|p^w_i - \tilde{N}^w_i|} \wedge \epsilon(n)\right) d\mathcal{H}^2(n) \\
&= 8\pi\Gamma(k_1, k_2, k_3)\Big(\sum_{i=1}^{k} |P^u_i - \tilde{N}^u_i| + \sum_{i=1}^{m} |P^w_i - \tilde{N}^w_i|\Big) \\
&\geq 8\pi\Gamma(k_1, k_2, k_3) L(u, w),
\end{aligned}$$

where $L(u, w)$ is the minimal connection of w and u, that is, the minimal connection between $\{P^u_i\}^k_{i=1} \cup \{N^w_i\}^m_{i=1}$ and $\{N^u_i\}^k_{i=1} \cup \{P^w_i\}^m_{i=1}$.

By the density result of [Bethuel 1990], the last equation is true for all $w, u \in H^1_\gamma(B, S^2)$ with $w - u \in H^1_0(B, S^2)$.

Now, taking $u = u_\lambda$ for $0 \le \lambda < 1$, let w be any map $H^1_\gamma(\Omega, S^2)$ with $u_\lambda - w \in H^1_0(B, S^2)$ with an arbitrary subdomain $B \subset \Omega$. By the minimality of u_λ, we have

$$E(u_\lambda, \Omega) + \lambda 8\pi\Gamma(k_1, k_2, k_3)L(u_\lambda, u_0) \le E(w; \Omega) + \lambda 8\pi\Gamma(k_1, k_2, k_3)L(w, u_0),$$

Moreover, we know

$$L(w, u_0) - L(u_\lambda, u_0) \le L(w, u_\lambda).$$

For $0 \le \lambda < 1$, we have

$$E(u_\lambda; B) \le \frac{1+\lambda}{1-\lambda} E(w; B)$$

for all $w \in H^1_{u_\lambda}(B, S^2)$. This proves our claim. □

Using Lemma 5.1 with an extension lemma in [Hardt et al. 1988], we have:

Proposition 5.2 (Caccioppoli's inequality). *For any* $0 \le \lambda < 1$, *let* u_λ *be a minimizer of* E_λ *in* $H^1_\gamma(\Omega, S^2)$. *Then for all* $x_0 \in \Omega$ *and* $R < dist(x_0, \partial\Omega)$, *we have*

$$\int_{B_{R/2}(x_0)} |\nabla u_\lambda|^2 \le C R^{-2} \int_{B_R(x_0)} |u_\lambda - u_{\lambda x_0, R}|^2 \, dx. \tag{5–1}$$

Next, we have:

Proposition 5.3 [Hong 2004]. *Let* $u \in H^1(\Omega, S^2)$ *be any weak solution of* (1–3) *and assume that u satisfies the Caccioppoli inequality* (5–1). *Then u is smooth in an open set* $\Omega_0 \subset \Omega$ *and* $\mathcal{H}^\beta(\Omega \backslash \Omega_0) = 0$ *for some positive* $\beta < 1$.

Proof of Theorem 1.3. It follows from Propositions 5.2 and 5.3. □

Finally, it seems that there exists no monotonicity formula for the minimizers of F_{u_0} in H^1. It is a challenging question whether one can establish the partial regularity of minimizers of F_{u_0} in $H^1(\Omega, S^2)$ for a given map $u_0 \in R^\infty_\gamma$.

Acknowledgment

The research of the author has been supported by the Australian Research Council. The partial research was started in November 2003 when the author participated in the Intensive Research Period on Geometric Analysis at the Research Center Ennio De Giorgi. The author would like to thank Professor Mariano Giaquinta for valuable discussions.

References

[Bethuel 1990] F. Bethuel, "A characterization of maps in $H^1(B^3, S^2)$ which can be approximated by smooth maps", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **7**:4 (1990), 269–286. MR 91f:58013 Zbl 0708.58004

[Bethuel and Brézis 1991] F. Bethuel and H. Brézis, "Regularity of minimizers of relaxed problems for harmonic maps.", *J. Funct. Anal.* **101**:1 (1991), 145–161. Zbl 0797.49034

[Bethuel et al. 1990] F. Bethuel, H. Brezis, and J.-M. Coron, "Relaxed energies for harmonic maps", pp. 37–52 in *Variational methods* (Paris, 1988), edited by H. Berestycki et al., Progr. Nonlinear Differential Equations Appl. **4**, Birkhäuser, Boston, 1990. MR 94a:58046 Zbl 0793.58011

[Brezis and Coron 1983] H. Brezis and J.-M. Coron, "Large solutions for harmonic maps in two dimensions", *Comm. Math. Phys.* **92**:2 (1983), 203–215. MR 85a:58022 Zbl 0532.58006

[Brezis et al. 1986] H. Brezis, J.-M. Coron, and E. H. Lieb, "Harmonic maps with defects", *Comm. Math. Phys.* **107**:4 (1986), 649–705. MR 88e:58023 Zbl 0608.58016

[Giaquinta 1983] M. Giaquinta, *Multiple integrals in the calculus of variations and nonlinear elliptic systems*, Annals of Mathematics Studies **105**, Princeton University Press, Princeton, NJ, 1983. MR 86b:49003 Zbl 0516.49003

[Giaquinta et al. 1989] M. Giaquinta, G. Modica, and J. Souček, "Cartesian currents and variational problems for mappings into spheres", *Ann. Sc. Norm. Super. Pisa* **16**:3 (1989), 393–485. Zbl 0713.49014

[Giaquinta et al. 1990] M. Giaquinta, G. Modica, and J. Souček, "Liquid crystals: Relaxed energies, dipoles, singular lines and singular points", *Ann. Sc. Norm. Super. Pisa* **17**:3 (1990), 415–437. Zbl 0760.49026

[Giaquinta et al. 1998] M. Giaquinta, G. Modica, and J. Souček, *Cartesian currents in the calculus of variations. II*, Ergebnisse der Mathematik **38**, Springer, Berlin, 1998. Variational integrals. MR 2000b:49001b Zbl 0914.49002

[Hardt et al. 1986] R. Hardt, D. Kinderlehrer, and F.-H. Lin, "Existence and partial regularity of static liquid crystal configurations", *Comm. Math. Phys.* **105**:4 (1986), 547–570. MR 88a:35207 Zbl 0611.35077

[Hardt et al. 1988] R. Hardt, D. Kinderlehrer, and F.-H. Lin, "Stable defects of minimizers of constrained variational principles", *Ann. Inst. Henri Poincaré, Analyse non linéaire* **5** (1988), 297–322. Zbl 0657.49018

[Hong 2004] M.-C. Hong, "Partial regularity of weak solutions of the liquid crystal equilibrium system", *Indiana Univ. Math. J.* **53**:5 (2004), 1401–1414. MR 2005k:35127 Zbl 02211012

[Isobe 1995] T. Isobe, "Energy gap phenomenon and the existence of infinitely many weakly harmonic maps for the Dirichlet problem", *J. Funct. Anal.* **129**:2 (1995), 243–267. MR 96c:58048 Zbl 0842.58013

[Pakzad 2001] M. R. Pakzad, "Existence of infinitely many weakly harmonic maps from a domain in $\mathbb{R}^n$ into S^2 for non-constant boundary data", *Calc. Var. Partial Differ. Equ.* **13**:1 (2001), 97–121. Zbl 0993.58008

[Rivière 1995] T. Rivière, "Everywhere discontinuous harmonic maps into spheres", *Acta Math.* **175**:2 (1995), 197–226. Zbl 0898.58011

Received June 8, 2006. Revised February 18, 2007.

MIN-CHUN HONG

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF QUEENSLAND
BRISBANE, QLD 4072
AUSTRALIA
hong@maths.uq.edu.au

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

SUR LE DÉPLOIEMENT DES FORMES BILINÉAIRES EN CARACTÉRISTIQUE 2

AHMED LAGHRIBI

This article deals with the standard splitting of bilinear forms in characteristic 2. The first part is devoted to the study of bilinear Pfister neighbors (the definition of such a bilinear form is slightly different from the classical definition of a Pfister neighbor quadratic form). In the second part, we introduce the degree invariant for bilinear forms and we prove that for any integer $d \geq 0$, the d-th power of the ideal of even dimensional bilinear forms coincides with the set of bilinear forms of degree $\geq d$ (this is a positive answer to the analogue of the degree conjecture for quadratic forms). In the third part, we classify good bilinear forms of height 2, and we give information on the possible dimensions of bilinear forms of height 2 which are not necessarily good.

1. Introduction

Le but de cet article est d'étendre la théorie de déploiement standard aux formes bilinéaires en caractéristique 2. Cette théorie a été introduite en premier par M. Knebusch dans les années soixante-dix dans le cas des formes quadratiques en caractéristique $\neq 2$ [Knebusch 1976 ; 1977]. On la connait plutôt sous le nom de la théorie de déploiement générique, puisque dans ce cas la suite de déploiement standard d'une forme quadratique reflète des informations liées au comportement de la forme sur les extensions du corps de base. Récemment, en caractéristique 2, Knebusch et Rehmann ont étudié le déploiement standard des formes quadratiques de radical de dimension ≤ 1, et ont montré sa généricité comme ce qui est le cas en caractéristique $\neq 2$ [Knebusch and Rehmann 2000]. Ceci ne se généralise pas au cas des formes quadratiques de radical de dimension ≥ 2 [Hoffmann and Laghribi 2004, example 8.15]. Pour ces dernières, le déploiement standard a été traité dans [Laghribi 2002b], et une autre notion de généricité a été introduite dans [Knebusch $\geq$ 2007].

MSC2000: 11E04, 11E81.
Keywords: symmetric bilinear form, function field of a bilinear form, standard splitting of bilinear forms, bilinear Pfister neighbors.

Pour la suite de cet article, on fixe F un corps commutatif de caractéristique 2, et l'expression “forme bilinéaire” signifiera “forme bilinéaire symétrique de dimension finie et de radical nul”.

A une forme bilinéaire B d'espace sous-jacent V, on associe une forme quadratique $\widetilde{B}$ définie par : $\widetilde{B}(v) = B(v, v)$ pour $v \in V$. Cette forme quadratique est totalement singulière[1] et est unique à isométrie près, on l'appelle la forme quadratique associée à B. Le corps de fonctions de B, qu'on note $F(B)$, est défini comme étant celui de $\widetilde{B}$. On note $\dim B$ (*resp.* B_{an}) la dimension de B (*resp.* la partie anisotrope de B).

La tour de déploiement standard d'une forme bilinéaire B non nulle est une suite $(B_i, F_i)_{0 \leq i \leq h}$ donnée par :

$$\begin{cases} F_0 = F \quad \text{et} \quad B_0 = B_{\mathrm{an}} \\ \text{Pour } n \geq 1 : \quad F_n = F_{n-1}(B_{n-1}) \quad \text{et} \quad B_n = ((B_{n-1})_{F_n})_{\mathrm{an}}. \end{cases}$$

La hauteur (standard) de B, qu'on note $\mathrm{h}(B)$, est le plus petit entier h vérifiant $\dim B_h \leq 1$. En plus de la hauteur, on montre qu'il existe une d-forme bilinéaire de Pfister π unique telle que $B_{\mathrm{h}(B)-1}$ soit semblable à une sous-forme de π de dimension $2^d - 1$ ou 2^d suivant que $\dim B$ est impaire ou paire. L'entier d s'appelle le degré de B et on le note $\deg(B)$ (voir section 4).

On traitera le déploiement standard des formes bilinéaires en parallèle avec ce qui a été fait pour les formes quadratiques. Plus particulièrement, on s'intéressera à l'invariant degré et au problème de classification par hauteur et degré. Nos méthodes sont propres aux formes bilinéaires en caractéristique 2. Pour les faire deux difficultés se sont posées. D'une part, l'utilisation de l'analogue du théorème de la sous-forme (théorème 3.5) dont la formulation est basée sur la notion de forme bilinéaire et forme totalement singulière associées, qui est moins forte que la condition de sous-forme (ou de domination) utilisée dans le cas des formes quadratiques. D'autre part, on manque pour les formes bilinéaires d'un objet analogue à l'algèbre de Clifford d'une forme quadratique.

Maintenant on détaille le contenu de notre travail. Pour garder l'autonomie de cet article, on rappelera dans la section 2 quelques notions de base sur les formes bilinéaires et quadratiques en caractéristique 2.

La section 3 sera consacrée aux formes bilinéaires voisines de Pfister et vient compléter des résultats établis récemment dans [Laghribi 2005, Section 5]. Dans la sous-section 3A, on donnera des généralités sur les formes bilinéaires voisines. La définition d'une telle forme utilise la notion de forme bilinéaire et forme totalement singulière associées. En terme de déploiement sur les corps de fonctions, on sait d'après [Laghribi 2005, Cor. 5.6] qu'une forme bilinéaire anisotrope B est une

[1] C'est-à-dire, une forme quadratique dont le radical coincide avec son espace sous-jacent.

voisine de Pfister si et seulement si il existe C et D des formes bilinéaires telles que $(C_{F(B)})_{\mathrm{an}} \simeq D_{F(B)}$ et que $\widetilde{B} \simeq \widetilde{C}$ ($\simeq$ désigne l'isométrie). Contrairement au cas des formes quadratiques voisines, on va donner un exemple où la forme bilinéaire $C \perp D$ est isotrope, et un autre où $(B_{F(B)})_{\mathrm{an}} \simeq D'_{F(B)}$ avec $B \perp D'$ isotrope (Exemple 3.11). De plus, on va voir qu'une forme bilinéaire anisotrope B peut être une voisine de Pfister sans que la forme $(B_{F(B)})_{\mathrm{an}}$ soit définie sur F (Exemple 3.12). On finira cette sous-section par un fait important affirmant que toute forme bilinéaire anisotrope devient une voisine de Pfister anisotrope après extension des scalaires à un corps convenable (proposition 3.13). Ce résultat nous sera très utile dans la sous-section 5C pour les formes bilinéaires de hauteur 2. La sous-section 3B sera consacrée à la classe des formes bilinéaires voisines anisotropes B pour lesquelles la forme $(B_{F(B)})_{\mathrm{an}}$ est définie sur F. On utilisera les formes de cette classe pour suggérer une définition de forme bilinéaire excellente.

Dans la section 4, on abordera l'invariant degré en montrant que l'ensemble des formes bilinéaires de degré $\geq d$ coincide avec l'idéal $I^d F$ pour tout entier $d \geq 0$ (théorème 4.5). Pour cela, on va se ramener au cas des formes quadratiques en associant à toute forme bilinéaire B la forme quadratique $B \otimes [1, t^{-1}]$ avec t une variable sur F. Ceci va permettre d'utiliser quelques résultats récents dus à Aravire et Baeza [2003].

La section 5 sera consacrée aux formes bilinéaires bonnes, c'est-à-dire, celles dont la forme bilinéaire de Pfister correspondant à l'avant-dernière forme de leurs tours de déploiement standard est définie sur F. Plus particulièrement, on classifiera les formes bilinéaires bonnes de hauteur 2 (proposition 5.9 et théorème 5.10). Pour cela, on utilisera entre autres une généralisation aux formes bilinéaires d'un résultat récent de Karpenko sur les dimensions des formes quadratiques de I^n en caractéristique $\neq 2$ (proposition 5.7). Notre classification fait paraître une classe de formes bilinéaires bonnes de hauteur 2 dont on n'a pas un analogue pour les formes quadratiques en caractéristique $\neq 2$ (Commentaire après le théorème 5.10 ; Exemple 5.11). Finalement dans la sous-section 5C on donnera des informations sur les éventuelles dimensions des formes bilinéaires de hauteur 2 non nécessairement bonnes (corollaire 5.20), et ce en utilisant une version raffinée de la décomposition de Witt d'une forme bilinéaire (proposition 5.15).

2. Quelques rappels

Soit φ une forme quadratique (*resp.* une forme bilinéaire) d'espace sous-jacent V. On dit que φ est isotrope s'il existe $v \in V - \{0\}$ tel que $\varphi(v) = 0$ (*resp.* $\varphi(v, v) = 0$). Dans le cas contraire, on dit que φ est anisotrope. On désigne par $D_F(\varphi)$ l'ensemble des scalaires de F^* représentés par φ (*resp.* l'ensemble des scalaires $\varphi(v, v) \in F^*$ avec $v \in V$).

Pour $n \geq 1$ un entier et B une forme bilinéaire (ou quadratique), on désigne par $n \times B$ la somme orthogonale de n copies de B.

Deux formes bilinéaires (ou quadratiques) B et B' sont dites semblables si $B \simeq \alpha B'$ pour un certain $\alpha \in F^*$.

Le radical d'une forme bilinéaire B (*resp.* d'une forme quadratique φ) d'espace sous-jacent V est l'espace $\{v \in V \mid B(v, V) = 0\}$ (*resp.* le radical de la forme bilinéaire B_φ associée à φ).

Une forme quadratique est dite non singulière si son radical est nul.

On sait qu'une forme quadratique non singulière (*resp.* totalement singulière) est isométrique à une somme orthogonale de formes de type $[a, b] = ax^2 + xy + by^2$ (*resp.* de type $[a] = ax^2$).

On note $W(F)$ (*resp.* $W_q(F)$) l'anneau de Witt des formes bilinéaires (*resp.* le groupe de Witt des formes quadratiques non singulières).

2A. *Décomposition de Witt.*

2A1. *Cas des formes quadratiques.* Toute forme quadratique φ se décompose, à isométrie près, comme suit :

$$\varphi \simeq \varphi_{\mathrm{an}} \perp i \times [0, 0] \perp j \times [0] \tag{1}$$

où φ_{an} est une forme anisotrope, qu'on appelle la partie anisotrope de φ [Hoffmann and Laghribi 2004]. L'entier i s'appelle l'indice de Witt de φ et on le note $i_W(\varphi)$.

Une forme quadratique non singulière φ est dite hyperbolique si $\dim \varphi = 2 i_W(\varphi)$.

2A2. *Cas des formes bilinéaires.* On note $\langle a_1 : b : a_2 \rangle$ la forme bilinéaire B dont l'espace sous-jacent a pour base $\{e_1, e_2\}$ qui satisfait les conditions $B(e_i, e_i) = a_i$ et $B(e_1, e_2) = b$. Un plan métabolique est une forme bilinéaire isométrique à $\langle a : 1 : 0 \rangle$ pour un certain $a \in F$.

Toute forme bilinéaire B se décompose de la manière suivante :

$$B \simeq M \perp B_{\mathrm{an}} \tag{2}$$

où M est une somme orthogonale de plans métaboliques, et B_{an} est une forme bilinéaire anisotrope. La forme B_{an} est unique à isométrie près [Milnor and Husemoller 1973; Knebusch 1970] ; on l'appelle la partie anisotrope de B. L'indice de Witt de B, qu'on note $i_W(B)$, est l'entier $\frac{1}{2} \dim M$.

Une forme bilinéaire B est dite métabolique si $\dim B = 2 i_W(B)$.

On renvoie à la proposition 5.15 pour une version raffinée de la décomposition donnée dans (2).

Deux formes quadratiques (*resp.* deux formes bilinéaires) B et B' sont dites équivalentes, qu'on note $B \sim B'$, lorsque les formes B_{an} et B'_{an} sont isométriques.

2B. ***Formes bilinéaires et formes quadratiques de Pfister.*** On note $\langle a_1, \dots, a_n\rangle$ la forme bilinéaire $\sum_{i=1}^n a_i x_i y_i$ avec $a_1, \dots, a_n \in F^*$. Une n-forme bilinéaire de Pfister est une forme isométrique à $\langle 1, a_1\rangle \otimes \cdots \otimes \langle 1, a_n\rangle$. On la note $\langle\langle a_1, \dots, a_n\rangle\rangle$. La partie pure d'une forme bilinéaire de Pfister B est l'unique forme B' vérifiant $B = \langle 1\rangle \perp B'$.

Une $(n+1)$-forme quadratique de Pfister est une forme isométrique à

$$\langle\langle a_1, \dots, a_n\rangle\rangle \otimes [1, b],$$

où $\otimes$ est l'action de module de $W(F)$ sur $W_q(F)$, et $a_1, \dots, a_n \in F^*$, $b \in F$.

Soit IF l'idéal de $W(F)$ formé des formes bilinéaires de dimension paire. On pose $I^n F = (IF)^n$ et $I_q^{n+1} F = I^n F \otimes W_q(F)$ pour tout $n \geq 0$ (avec $I^0 F = W(F)$).

On sait que $I^n F$ (*resp.* $I_q^{n+1} F$) est engendré additivement par les n-formes bilinéaires de Pfister (*resp.* les $(n+1)$-formes quadratiques de Pfister).

Une n-forme quadratique de Pfister est dite de degré n [Hoffmann and Laghribi 2004; Laghribi 2002b]. Notre définition de degré diffère de celle adoptée par Aravire et Baeza [2003]. La notre étend la définition de degré en caractéristique $\neq 2$.

2C. ***Formes quadratiques voisines.*** Soient φ et φ' deux formes quadratiques d'espaces sous-jacent respectifs V et V'. On dit que φ est dominée par φ', qu'on note $\varphi \prec \varphi'$, s'il existe une application linéaire injective $\sigma : V \longrightarrow V'$ telle que $\varphi'(\sigma(v)) = \varphi(v)$ pour tout $v \in V$. On renvoit à [Hoffmann and Laghribi 2004, lem. 3.1] pour une description équivalente à cette définition. Notons que la relation de domination n'est autre que la relation de sous-forme lorsque les formes φ et φ' sont non singulières ou totalement singulières

Une forme quadratique φ est dite voisine d'une forme de Pfister π si $2 \dim \varphi > \dim \pi$ et $a\varphi \prec \pi$ pour un certain $a \in F^*$. Lorsque φ est voisine de π, alors π est unique et pour toute extension K/F, la forme φ_K est isotrope si et seulement si π_K est isotrope.

On sait qu'une forme quadratique voisine ne peut être totalement singulière. Mais pour les formes totalement singulières on a aussi la notion de forme voisine (Définition 3.7).

3. Les formes bilinéaires voisines

3A. ***Généralités sur les formes bilinéaires voisines.*** La notion de forme bilinéaire et forme totalement singulière associées va jouer un rôle essentiel dans cet article. Le lemme qui suit donne une définition équivalente à cette notion :

Lemme 3.1. *Une forme quadratique totalement singulière φ est associée à une forme bilinéaire B si et seulement si* $\dim B = \dim \varphi$ *et* $D_F(B) = D_F(\varphi)$.

Démonstration. D'après [Laghribi 2004a, lem. 2.1], on sait que deux formes quadratiques totalement singulières sont isométriques si elles sont de même dimension et représentent les mêmes scalaires de F^*. □

Remarque 3.2. (1) Notons qu'une forme bilinéaire B est isotrope si et seulement si $\widetilde{B}$ est isotrope.
(2) La correspondance $B \mapsto \widetilde{B}$ est compatible avec la somme orthogonale et la multiplication par des scalaires de F^*.

Notations 3.3. Pour φ une forme quadratique totalement singulière, on notera $\mathcal{A}(\varphi)$ l'ensemble de toutes les formes bilinéaires associées à φ.

Définition 3.4. Une forme bilinéaire B est dite une sous-forme d'une autre forme C si $C \simeq B \perp B'$ pour une certaine forme bilinéaire B'.

La définition d'une forme bilinéaire voisine est motivée par l'analogue du théorème de la sous-forme dont voici la formulation :

Théorème 3.5 [Laghribi 2005, prop. 1.1]. *Soient B et C deux formes bilinéaires anisotropes telles que B devienne métabolique sur $F(C)$. Alors, pour tout $\alpha \in D_F(C)D_F(B)$, il existe B' une sous-forme de αB telle que $B' \in \mathcal{A}(\widetilde{C})$. En particulier*, $\dim C \leq \dim B$.

Définition 3.6 [Laghribi 2005, section 5]. Une forme bilinéaire B est dite voisine d'une forme bilinéaire de Pfister π si $2 \dim B > \dim \pi$ et s'il existe $B' \in \mathcal{A}(\widetilde{B})$ semblable à une sous-forme de π.

Dans le cas des formes quadratiques totalement singulières, les notions de formes de Pfister et leurs voisines se définissent comme suit :

Définition 3.7. (1) Une forme totalement singulière est une quasi n-forme de Pfister si elle est associée à une n-forme bilinéaire de Pfister.
(2) Une forme totalement singulière φ est dite une quasi-voisine de Pfister s'il existe une quasi-forme de Pfister π tels que $2 \dim \varphi > \dim \pi$ et $a\varphi \prec \pi$ pour un certain $a \in F^*$.

On renvoie à [Hoffmann and Laghribi 2004] et [Laghribi 2004a] pour plus de détails sur les formes quasi-voisines et leurs déploiements standard.

Les formes bilinéaires voisines et les formes quadratiques quasi-voisines se correspondent mutuellement comme le montre la proposition suivante :

Proposition 3.8. *Une forme bilinéaire anisotrope B est une voisine de Pfister si et seulement si $\widetilde{B}$ est une quasi-voisine de Pfister.*

Démonstration. Soit B une forme bilinéaire anisotrope.

Supposons que B soit voisine d'une forme bilinéaire de Pfister π. Alors on a $2 \dim B > \dim \pi$ et il existe $B' \in \mathcal{A}(\widetilde{B})$ qui est semblable à une sous-forme de π.

Puisque $\widetilde{B} \simeq \widetilde{B}'$, la forme $\widetilde{B}$ est semblable à une sous-forme de $\widetilde{\pi}$. Ainsi, $\widetilde{B}$ est une quasi-voisine de $\widetilde{\pi}$.

Réciproquement, supposons maintenant que $\widetilde{B}$ soit quasi-voisine d'une quasi-forme de Pfister τ. Soit C une forme bilinéaire de Pfister telle que $\widetilde{C} \simeq \tau$. Puisque $\tau_{F(B)}$ est isotrope, la forme $C_{F(B)}$ est isotrope et donc elle est métabolique [Laghribi 2005, prop. 3.3]. Par le théorème 3.5, il existe $B' \in \mathcal{A}(\widetilde{B})$ qui est semblable à une sous-forme de C. Puisque $2 \dim B > \dim \tau = \dim C$, la forme B est bien une voisine de C. □

On se servira souvent de la proposition suivante qui montre qu'une forme bilinéaire anisotrope est une voisine de Pfister lorsqu'elle devient métabolique sur le corps de fonctions d'une autre forme bilinéaire de dimension suffisamment grande :

Proposition 3.9 [Laghribi 2005, cor. 5.4]. *Soient B et C deux formes bilinéaires anisotropes. Si $B_{F(C)}$ est métabolique et $2 \dim C > \dim B$, alors B est semblable à une forme bilinéaire de Pfister π, et toute forme bilinéaire $B' \in \mathcal{A}(\widetilde{C})$ est voisine de π. En particulier, C est voisine de π.*

Comme dans le cas des formes quadratiques voisines ou quasi-voisines, les formes bilinéaires voisines vérifient certaines propriétés classiques :

Proposition 3.10 [Laghribi 2005, prop. 5.2, 5.3, cor. 5.6]. *Soient B et C deux formes bilinéaires anisotropes avec C une forme bilinéaire de Pfister.*
(1) *Si B est voisine de C, alors pour toute extension K/F les formes B_K et C_K sont simultanément isotropes ou anisotropes.*
(2) *B est voisine de C si et seulement si $2 \dim B > \dim C$ et $C_{F(B)}$ est isotrope.*
(3) *B est une voisine d'une forme bilinéaire de Pfister si et seulement si il existe $B' \in \mathcal{A}(\widetilde{B})$ telle que la forme $(B'_{F(B)})_{an}$ soit définie sur F.*

Cependant, d'autres propriétés sur les formes quadratiques voisines ou quasi-voisines ne se généralisent pas aux formes bilinéaires voisines. Par exemple, une forme bilinéaire peut être voisine de deux formes bilinéaires de Pfister non isométriques. De plus, contrairement à un résultat classique de Fitzgerald [1981, th. 1.6], l'exemple suivant illustre un cas d'une forme bilinéaire voisine anisotrope B et d'une forme $C \in \mathcal{A}(\widetilde{B})$ telles que $(C_{F(B)})_{\mathrm{an}} \simeq D_{F(B)}$ pour une certaine forme D mais que $C \perp D$ est isotrope.

Exemple 3.11. Soient $x_1, \ldots, x_d, u, v$ des variables sur un corps F_0 de caractéristique 2 ($d \geq 1$), $F = F_0(x_i, u, v)$ et $R = \langle\langle x_1, \ldots, x_d \rangle\rangle$. Soient

$$B = \langle 1, 1+u, v, uv \rangle \otimes R, \quad C = \langle 1, u, u+v, uv \rangle \otimes R$$

et $\pi = \langle\langle u, v \rangle\rangle \otimes R$ qui sont des formes anisotropes. Alors :
(1) B est une voisine de π.
(2) $(B_{F(B)})_{\mathrm{an}} \simeq (\langle u, 1+u \rangle \otimes R)_{F(B)}$ et $B \perp \langle u, 1+u \rangle \otimes R$ est isotrope.

(3) $C \in \mathscr{A}(\widetilde{B})$.
(4) $(C_{F(B)})_{\text{an}} \simeq (\langle v, u+v\rangle \otimes R)_{F(B)}$ et $C \perp \langle v, u+v\rangle \otimes R$ est isotrope.

Démonstration. Puisque les formes $\langle 1, 1+u, v, uv\rangle$ et $\langle 1, u, u+v, uv\rangle$ sont associées à la forme quadratique $[1] \perp [u] \perp [v] \perp [uv]$, on déduit que $\widetilde{B} \simeq \widetilde{\pi} \simeq \widetilde{C}$. Ainsi, B est voisine de π et $C \in \mathscr{A}(\widetilde{B})$. On a $B \sim \pi \perp \langle u, 1+u\rangle \otimes R$ et $C \sim \pi \perp \langle v, u+v\rangle \otimes R$. Puisque $\pi_{F(B)}$ est isotrope et les formes $(\langle u, 1+u\rangle \otimes R)_{F(B)}$ et $(\langle v, u+v\rangle \otimes R)_{F(B)}$ sont anisotropes (par raison de dimension et le théorème 3.5), on déduit que $(B_{F(B)})_{\text{an}} \simeq (\langle u, 1+u\rangle \otimes R)_{F(B)}$ et $(C_{F(B)})_{\text{an}} \simeq (\langle v, u+v\rangle \otimes R)_{F(B)}$. □

L'exemple qui va suivre montre qu'une forme bilinéaire anisotrope B peut être une voisine de Pfister sans que la forme $(B_{F(B)})_{\text{an}}$ soit définie sur F :

Exemple 3.12. Soient x, y, z des variables sur un corps F_0 de caractéristique 2, et $F = F_0(x, y, z)$. Soit $B = \langle x, y, xy, 1+x, z, (1+x)z\rangle$. Alors, B est une voisine de Pfister mais la forme $(B_{F(B)})_{\text{an}}$ n'est pas définie sur F.

Démonstration. Puisque $[x] \perp [1+x] \simeq [1] \perp [x]$ et $[z] \perp [z(1+x)] \simeq [z] \perp [xz]$, on obtient que B est voisine de $\langle\langle x, y, z\rangle\rangle$. On a nécessairement $\dim(B_{F(B)})_{\text{an}} = 4$ puisque $B_{F(B)}$ ne peut être métabolique. Ainsi, B est de hauteur et de degré 2. De plus, par le théorème 5.10 B ne peut être bonne et donc la forme $(B_{F(B)})_{\text{an}}$ n'est pas définie sur F. □

On finit cette sous-section par un résultat qui montre qu'une forme bilinéaire anisotrope devient une voisine de Pfister anisotrope après extension des scalaires à un corps convenable. L'ingrédient essentiel qu'on utilise est la notion de degré normique d'une forme totalement singulière. On renvoie à [Hoffmann and Laghribi 2004, section 8] pour plus de détails sur cet invariant et certaines de ses applications. Rappelons tout de même que le corps normique d'une forme quadratique totalement singulière non nulle φ, qu'on note $N_F(\varphi)$, est défini par $N_F(\varphi) = F^2(ab \mid a, b \in D_F(\varphi))$. Le degré $[N_F(\varphi) : F^2]$ s'appelle le degré normique de φ, et on le note $\operatorname{ndeg}_F(\varphi)$.

Proposition 3.13. *Soient $n \geq 1$ un entier et B une forme bilinéaire anisotrope telle que* $\dim B \in]2^n, 2^{n+1}]$. *Posons* $\operatorname{ndeg}_F(\widetilde{B}) = 2^l$.
(1) *On a $l \geq n+1$, et B est une voisine de Pfister si et seulement si $l = n+1$.*
(2) *Si $l > n+1$, alors il existe une suite de formes bilinéaires $\pi_{l-n-2} \subset \cdots \subset \pi_0$ telle que chaque π_i soit une $(l-i)$-forme bilinéaire de Pfister et B_K soit une voisine de Pfister anisotrope, où $K = F(\pi_0)\cdots(\pi_{l-n-2})$.*

Démonstration. (1) L'inégalité $l \geq n+1$ provient de [Hoffmann and Laghribi 2004, prop. 8.6]. De plus, par [Hoffmann and Laghribi 2004, prop. 8.9] on a que $\widetilde{B}$ est une quasi-voisine de Pfister si et seulement si $l = n+1$, et par la proposition 3.8 ceci équivaut à dire que B est une voisine de Pfister.

(2) On reprend la même preuve de [Laghribi 2004b, prop. 1.9] appliquée à la forme $\widetilde{B}$, et on utilise la proposition 3.8. □

3B. *Les formes bilinéaires voisines strictes.*

Définition 3.14. Une forme bilinéaire anisotrope B est dite une voisine stricte s'il existe une forme bilinéaire anisotrope C telle que $(B_{F(B)})_{\mathrm{an}} \simeq C_{F(B)}$.

Supposons que $B \perp C$ soit anisotrope.

Lemme 3.15. *Soient B, C, C' des formes bilinéaires anisotropes telles que*

$$(B_{F(B)})_{an} \simeq C_{F(B)} \simeq C'_{F(B)}.$$

Si $B \perp C$ est anisotrope, alors $B \perp C'$ est aussi anisotrope.

Démonstration. Par la proposition 3.9, la forme $B \perp C$ (*resp.* $(B \perp C')_{\mathrm{an}}$) est semblable à une forme bilinéaire de Pfister π_1 (*resp.* π_2) dont B est voisine. Puisque $\widetilde{B}$ est semblable à une sous-forme de $\widetilde{\pi}_2$ et que B est isotrope sur $F(\pi_1)$, on déduit que π_2 est isotrope sur $F(\pi_1)$ et donc elle est métabolique sur $F(\pi_1)$. En particulier, $\dim \pi_1 \leq \dim \pi_2$. Comme $\dim \pi_2 \leq \dim \pi_1$, on a nécessairement $\dim(B \perp C) = \dim(B \perp C')_{\mathrm{an}}$. Ainsi, $B \perp C'$ est anisotrope. □

Soit B une forme bilinéaire voisine stricte et C une autre forme bilinéaire telle $(B_{F(B)})_{\mathrm{an}} \simeq C_{F(B)}$. Par le lemme 3.15, B est de l'un des deux types suivants qui s'excluent mutuellement :

Type I : si $B \perp C$ est isotrope.

Type II : si $B \perp C$ est anisotrope.

Voici une propriété sur les formes voisines strictes de type II qui les rapprochent des formes quadratiques voisines :

Proposition 3.16. *Soient B et C des formes bilinéaires anisotropes. Si $(B_{F(B)})_{an} \simeq C_{F(B)}$ et $B \perp C$ est anisotrope, alors $B \perp C$ est semblable à une forme bilinéaire de Pfister et la forme C est unique.*

Démonstration. Soit $n \geq 1$ un entier tel que $\dim B \in]2^{n-1}, 2^n]$. Puisque $\dim B > \dim C$ et $(B \perp C)_{F(B)} \sim 0$, alors on obtient par la proposition 3.9 que $B \perp C$ est semblable d'une m-forme bilinéaire de Pfister dont B est voisine. Ainsi, $n = m$ et $\dim C < 2^{n-1} < \dim B$. Si C' est une forme bilinéaire telle que $(B_{F(B)})_{\mathrm{an}} \simeq C'_{F(B)}$, alors par le lemme 3.15 $B \perp C'$ est anisotrope. Comme pour la forme C, on a $\dim C' < 2^{n-1} < \dim B$. Puisque $(C \perp C')_{F(B)} \sim 0$ et $2 \dim B > 2^n > \dim C + \dim C'$, on obtient par la proposition 3.9 que $C \simeq C'$. □

Cette proposition motive la définition suivante :

Définition 3.17. Soient B et C deux formes bilinéaires anisotropes telles que $(B_{F(B)})_{\mathrm{an}} \simeq C_{F(B)}$ et $B \perp C$ est anisotrope. La forme C est appelée la forme complémentaire de B.

Lemme 3.18. *Soit B une forme bilinéaire anisotrope qui est une voisine stricte de type II et de forme complémentaire C. Si K est une extension de F telles que B_K et C_K soient anisotropes, alors $(B_{K(B)})_{an} \simeq C_{K(B)}$.*

Démonstration. La condition $(B_{F(B)})_{\mathrm{an}} \simeq C_{F(B)}$ donne $B_{K(B)} \sim C_{K(B)}$. Comme B est de forme complémentaire C, on obtient que $\dim B > 2^n > \dim C$ pour un certain entier $n \geq 1$. Comme B_K et C_K sont anisotropes, on a par [Hoffmann and Laghribi 2006, th. 1.1] que $C_{K(B)}$ est anisotrope et donc $(B_{K(B)})_{\mathrm{an}} \simeq C_{K(B)}$. □

Définition 3.19. Soient B et C des formes bilinéaires anisotropes telles que $\dim B > \dim C$. Un couple de formes bilinéaires (B', C') est dit lié au couple (B, C) s'il existe une forme bilinéaire η telle que : $B \simeq B' \perp \eta$, $C \simeq C' \perp \eta$ et $(B \perp C)_{\mathrm{an}} \simeq B' \perp C'$. Dans ce cas, on note $(B, C) \dashrightarrow (B', C')$.

Le résultat suivant montre que de toute forme voisine stricte on peut se ramener au cas d'une forme voisine stricte de type II :

Proposition 3.20. *Soient B et C des formes bilinéaires anisotropes telles que $\dim B > \dim C$. Soient B' et C' des formes bilinéaires telles que $(B, C) \dashrightarrow (B', C')$.*
(1) *On a équivalence entre les assertions suivantes*:
(i) $(B_{F(B)})_{an} \simeq C_{F(B)}$.
(ii) $(B'_{F(B')})_{an} \simeq C'_{F(B')}$, *$B'_{F(B)}$ est isotrope et $C_{F(B)}$ est anisotrope.*
(2) *Si l'une des conditions équivalentes de* (1) *est vérifiée, alors B et B' sont voisines de la même forme bilinéaire de Pfister, et B' est une voisine stricte de type II.*

Démonstration. Soient B' et C' des formes bilinéaires telles que $(B,C) \dashrightarrow (B',C')$.

(1) (i) $\Longrightarrow$ (ii) Supposons que $(B_{F(B)})_{\mathrm{an}} \simeq C_{F(B)}$. Alors, $C_{F(B)}$ est anisotrope et $(B' \perp C')_{F(B)} \sim 0$. Puisque $B' \perp C'$ est anisotrope et

$$\dim(B' \perp C') \leq \dim(B \perp C) < 2 \dim B,$$

on déduit par la proposition 3.9 que $B' \perp C'$ est semblable à une n-forme de Pfister dont B est voisine. En particulier, $B'_{F(B)}$ est isotrope puisque $\dim B' > \dim C'$ (car $\dim B > \dim C$). Comme $\dim B' > 2^{n-1} > \dim C'$, on déduit par [Hoffmann and Laghribi 2006, th. 1.1] que $C'_{F(B')}$ est anisotrope. Puisque $(B' \perp C')_{F(B')} \sim 0$, on obtient $(B'_{F(B')})_{\mathrm{an}} \simeq C'_{F(B')}$.

(ii) $\Longrightarrow$ (i) Puisque $(B'_{F(B')})_{\mathrm{an}} \simeq C'_{F(B')}$, la proposition 3.9 implique que $B' \perp C'$ est semblable à une forme bilinéaire de Pfister. De l'isotropie de $B'_{F(B)}$ on déduit que $(B' \perp C')_{F(B)} \sim 0$. Ainsi, $(B \perp C)_{F(B)} \sim 0$. Comme $C_{F(B)}$ est anisotrope, on a bien $(B_{F(B)})_{\mathrm{an}} \simeq C_{F(B)}$.

(2) Si on a l'une des conditions équivalentes de (1), il est clair que B' est une voisine stricte de type II. Comme $\dim B \geq \dim B'$ et $B'_{F(B)}$ est isotrope, on obtient par la proposition 3.10 que B et B' sont voisines d'une même forme bilinéaire de Pfister. □

En vue de l'introduction de la notion de forme complémentaire dans le cas des formes voisines strictes de type II, on suggère la définition suivante d'une forme bilinéaire excellente :

Définition 3.21. Une forme bilinéaire anisotrope B est dite excellente si $\dim B \leq 1$ ou bien $\dim B > 1$ et B est une voisine stricte de type II et de forme complémentaire une forme excellente.

Voici une description de la tour de déploiement standard d'une forme bilinéaire excellente :

Proposition 3.22. *Soit B une forme bilinéaire anisotrope excellente de dimension ≥ 2 et de tour de déploiement standard $(B_i, F_i)_{0\leq i\leq \mathrm{h}(B)}$. Alors, $\mathrm{h}(B) \geq 1$ et il existe une suite de formes bilinéaires $(C_i)_{0\leq i\leq \mathrm{h}(B)}$ telle que*:
(1) *$B_i \simeq (C_i)_{F_i}$ pour tout $i \in \{0, \dots, \mathrm{h}(B)\}$.*
(2) *C_i est une voisine stricte de type II et de forme complémentaire C_{i+1} pour tout $i \in \{0, \dots, \mathrm{h}(B) - 1\}$.*

Démonstration. Il est évident que $h = \mathrm{h}(B) \geq 1$. Supposons que B soit excellente. Alors, il existe une suite $(C_i)_{0\leq i\leq k}$ de formes bilinéaires telles que $C_0 = B$, C_k est la forme nulle, et C_i soit une voisine stricte de type II et de forme complémentaire C_{i+1}. Supposons qu'on ait $B_i \simeq (C_i)_{F_i}$ pour un certain $i < \mathrm{h}(B)$. Alors, $F_{i+1} = F_i(C_i)$ et on a

$$(B_i)_{F_{i+1}} \sim B_{i+1} \sim (C_i)_{F_i(C_i)} \sim (C_{i+1})_{F_{i+1}}. \tag{3}$$

Comme C_i est une voisine de forme complémentaire C_{i+1} et que C_i est anisotrope sur F_i, alors la forme C_{i+1} est aussi anisotrope sur F_i. On déduit par le lemme 3.18 que $((C_i)_{F_i(C_i)})_{\mathrm{an}} \simeq (C_{i+1})_{F_{i+1}}$, et la relation (3) implique que $B_{i+1} \simeq (C_{i+1})_{F_{i+1}}$.

Ainsi de suite, on aboutit à $B_{h-1} \simeq (C_{h-1})_{F_{h-1}}$ et donc la forme C_h est forcèment nulle. En particulier, $k = \mathrm{h}(B)$. □

4. L'invariant degré pour les formes bilinéaires

Pour introduire le degré d'une forme bilinéaire, on se basera sur le théorème suivant qui classifie les formes bilinéaires de hauteur 1 :

Théorème 4.1. *Soit B une forme bilinéaire anisotrope. Alors, B est de hauteur 1 si et seulement si B est semblable à une forme bilinéaire de Pfister ou est semblable à la partie pure d'une forme bilinéaire de Pfister.*

Démonstration. Le théorème a été prouvé dans [Laghribi 2005, cor. 5.5] lorsque $\dim B$ est paire. Supposons que $\dim B$ soit impaire. Alors, par hypothèse on a $(B_{F(B)})_{\mathrm{an}} \simeq \langle\alpha\rangle_{F(B)}$ avec $\alpha = \det B$. Par conséquent, $(B \perp \langle\alpha\rangle)_{F(B)} \sim 0$. On a $B \perp \langle\alpha\rangle \not\sim 0$, et le théoème 3.5 donne $\dim B \leq \dim(B \perp \langle\alpha\rangle)_{\mathrm{an}}$. Ainsi, $B \perp \langle\alpha\rangle$ est anisotrope. Par la proposition 3.9, $B \perp \langle\alpha\rangle \simeq \beta\pi$ avec π une forme bilinéaire de Pfister et $\beta \in F^*$. Par la multiplicativité d'une forme bilinéaire de Pfister [Baeza 1978, cor. 2.16, page 101], et l'unicité de la partie pure d'une telle forme (cor. 2.18, *ibid.*), il est clair que $\alpha B \simeq \pi'$. □

Corollaire 4.2. *Soit B une forme bilinéaire non nulle de tour de déploiement standard $(F_i, B_i)_{0 \leq i \leq h}$ avec $h = \mathrm{h}(B)$. Alors, il existe une unique forme bilinéaire de Pfister π telle que B_{h-1} soit semblable à π ou semblable à la partie pure de π suivant que* $\dim B$ *est paire ou impaire.*

Démonstration. On utilise le théorème 4.1 et le fait que B_{h-1} est de hauteur 1, ainsi que la multiplicativité et l'unicité de la partie pure d'une forme de Pfister. □

Définition 4.3. Soit B une forme bilinéaire non nulle de tour de déploiement standard $(F_i, B_i)_{0 \leq i \leq h}$ avec $h = \mathrm{h}(B) \geq 1$.
(1) Le corps F_{h-1} s'appelle le corps dominant de B.
(2) La forme dominante de B est l'unique forme bilinéaire de Pfister correspondant à B_{h-1} au sens du corollaire 4.2.
(3) Si $\dim B$ est paire, le degré de B est l'entier d tel que $\dim B_{h-1} = 2^d$. Si $\dim B$ est impaire, on dit que B est de degré 0. Dans les deux cas, on désigne par $\deg(B)$ le degré de B.

Notations 4.4. Pour tout entier $n \geq 0$, on désigne par :
(1) $J_n^b(F)$ l'ensemble des formes bilinéaires de degré $\geq n$.
(2) $\overline{I^n} F$ le quotient $I^n F / I^{n+1} F$.
(3) $\overline{I^n}(K/F)$ le noyau de l'homomorphisme $\overline{I^n} F \longrightarrow \overline{I^n} K$ induit par l'inclusion $F \subset K$.

Le résultat principal de cette section est le théorème suivant :

Théorème 4.5. *Pour tout entier $n \geq 0$, on a $I^n F = J_n^b(F)$.*

On introduit quelques résultats préliminaires nécessaires pour la preuve de ce théorème.

Lemme 4.6. *Soient t une variable sur F, B une forme bilinéaire sur F et $\varphi = B \otimes [1, t^{-1}]$. Alors:*
(1) *B est isotrope si et seulement si φ est isotrope.*
(2) *$i_W(\varphi) = 2 i_W(B)$. En particulier, B est métabolique si et seulement si φ est hyperbolique.*

Démonstration. (1) Si φ est anisotrope, alors B est aussi anisotrope. Supposons que B soit anisotrope. On pose $B \simeq \langle a_1\rangle \perp \cdots \perp \langle a_n\rangle$ pour $a_1, \cdots, a_n \in F^*$ convenables. On a $\varphi \simeq a_1[1, t^{-1}] \perp \cdots \perp a_n[1, t^{-1}]$. Supposons que φ soit isotrope, et soit $(p_1, q_1, \ldots, p_n, q_n) \in F(t)^{2n} - \{0\}$ tel que :

$$\sum_{i=1}^{n} a_i(p_i^2 + p_i q_i + t^{-1} q_i^2) = 0. \tag{4}$$

Sans perdre de généralités, on peut supposer que $p_1, q_1, \ldots, p_n, q_n$ sont des polynômes non tous divisibles par t. On multiplie (4) par t, on substitue 0 à t, et on utilise l'anisotropie de B pour déduire que les polynômes q_i sont tous divisibles par t. On substitue de nouveau 0 à t et on utilise l'anisotropie de B pour déduire cette fois-ci que t divise tous les polynômes p_i, une contradiction.

(2) Soit $B \simeq M \perp B_{\mathrm{an}}$ la décomposition de Witt de B. On a $\varphi \simeq M \otimes [1, t^{-1}] \perp B_{\mathrm{an}} \otimes [1, t^{-1}]$ et la forme $M \otimes [1, t^{-1}]$ est hyperbolique. Par (1) on a que $B_{\mathrm{an}} \otimes [1, t^{-1}]$ est anisotrope. Ainsi, $i_W(\varphi) = \dim M = 2i_W(B)$. □

Lemme 4.7. *Soient t une variable sur F, B une forme bilinéaire non nulle de degré d. Alors,* $\deg(B \otimes [1, t^{-1}]) \leq d + 1$.

Démonstration. Posons $K = F(t)$. Soient $(F_i, B_i)_{0\leq i\leq h}$ et $(K_j, \varphi_j)_{0\leq j\leq k}$ les tours de déploiement standard respectives de B et φ, avec $h = \mathrm{h}(B)$ et $k = \mathrm{h}(\varphi)$. On a donc $\dim(B_{F_{h-1}})_{\mathrm{an}} = 2^d$. Par le lemme 4.6, $\dim(\varphi_{K\cdot F_{h-1}})_{\mathrm{an}} = 2^{d+1}$. Puisque $(K_j, \varphi_j)_{0\leq j\leq k}$ est la tour de déploiement générique de φ [Knebusch and Rehmann 2000], il existe $1 \leq j \leq k-1$ tel que $\dim \varphi_j = 2^{d+1}$. Ainsi, $\deg(\varphi) \leq d + 1$. □

On donne l'analogue du Hauptsatz d'Arason-Pfister pour les formes bilinéaires. Baeza a prouvé le même résultat dans le cas des formes quadratiques en caractéristique 2 [Baeza 1973] :

Lemme 4.8. *Soit $n \geq 0$ un entier et B une forme bilinéaire non métabolique. Si $B \in I^n F$, alors* $\dim B_{an} \geq 2^n$. *Si* $\dim B_{an} = 2^n$, *alors B_{an} est semblable à une n-forme bilinéaire de Pfister.*

Démonstration. Sans perdre de généralités, on peut supposer que B est anisotrope. Soit t une variable sur F et $\varphi = B \otimes [1, t^{-1}]$. On a $\varphi \in I_q^{n+1} F(t)$. Par le lemme 4.6, φ est anisotrope. Par [Baeza 1973] $\dim \varphi \geq 2^{n+1}$, i.e., $\dim B \geq 2^n$. Si $\dim B = 2^n$, alors $B_{F(B)} \sim 0$ puisque $B_{F(B)} \in I^n F(B)$. Par le théorème 4.1, B est semblable à une n-forme bilinéaire de Pfister. □

On aura besoin du théorème de la sous-forme dans le cas des formes quadratiques :

Théorème 4.9 [Hoffmann and Laghribi 2004, th. 4.2]. *Soient φ et φ' deux formes quadratiques anisotropes telles que φ' soit non singulière et devienne hyperbolique sur $F(\varphi)$. Alors, $\varphi \prec \alpha\varphi'$ pour tout scalaire $\alpha \in D_F(\varphi)D_F(\varphi')$.*

Comme on l'a évoqué dans l'introduction, l'analogue du théorème 4.5 pour les formes quadratiques en caractéristique 2 est dû à Aravire et Baeza [2003] :

Théorème 4.10. *Pour tout entier $n \geq 1$, on a que $I_q^n F$ est l'ensemble des formes quadratiques non singulières de degré $\geq n$.*

Ce théorème 4.10 permet de déduire le corollaire suivant :

Corollaire 4.11. *Soient $n \geq 0$ un entier, $\varphi \in I_q^{n+1} F$ et ψ une forme quadratique totalement singulière anisotrope de dimension $> 2^n$. Si $\varphi_{F(\psi)} \in I_q^{n+2} F(\psi)$, alors $\varphi \in I_q^{n+2} F$.*

Démonstration. Supposons que $\varphi \notin I_q^{n+2} F$. Par le théorème 4.10, on déduit que φ est de degré $n+1$. Soient $(F_i, \varphi_i)_{0 \leq i \leq h}$ sa tour de déploiement standard et π sa forme dominante avec $h = \mathrm{h}(\varphi)$. Puisque $\pi_{F_{h-1}(\psi)} \in I^{n+2} F_{h-1}(\psi)$, on obtient par le lemme 4.8 que $\pi_{F_{h-1}(\psi)} \sim 0$. Comme ψ est totalement singulière de dimension $> 2^n$, on obtient par le théorème 4.9 que $\pi \sim 0$, une contradiction. $\square$

Lemme 4.12. *Soient t une variable sur F, $B \in I^n F$ une forme bilinéaire telle que $B \otimes [1, t^{-1}] \in I_q^{n+2} F(t)$. Alors, $B \in I^{n+1} F$.*

Démonstration. Posons $B = \pi_1 \perp \cdots \perp \pi_m$ avec $\pi_1, \cdots, \pi_m$ des formes semblables à des n-formes bilinéaires de Pfister anisotropes. On procède par induction sur m. Si $m \leq 1$, alors le lemme 4.8 implique que $B \otimes [1, t^{-1}]$ est hyperbolique, et donc B est métabolique par le lemme 4.6. Supposons que $m \geq 2$. Sur le corps $L = F(\pi_m)$ et par induction sur m, on a que $B_L \in I^{n+1} L$. Ainsi, $B + I^{n+1} F \in \overline{I^n}(L/F)$. Par un résultat de Aravire et Baeza [2003], on déduit que $B \perp \pi \in I^{n+1} F$ pour une n-forme bilinéaire de Pfister convenable. Puisque $B \otimes [1, t^{-1}] \in I_q^{n+2} F(t)$, on obtient que $\pi \otimes [1, t^{-1}] \in I^{n+2} F(t)$. Les lemmes 4.6 et 4.8 impliquent que $\pi \sim 0$. Ainsi, $B \in I^{n+1} F$. $\square$

Le résultat suivant étend un calcul fait auparavant par Aravire et Baeza uniquement dans le cas du corps de fonctions d'une forme bilinéaire de Pfister [Aravire and Baeza 2003] :

Proposition 4.13. *Soit $n \geq 1$ un entier et ψ une forme quadratique totalement singulière telle que* $\dim \psi_{an} > 2^n$. *Alors, $\overline{I^n}(F(\psi)/F) = \{0\}$.*

Démonstration. On sait que si ψ est isotrope, alors $F(\psi)$ est une extension transcendante pure de $F(\psi_{\mathrm{an}})$. Donc, sans perdre de généralités, on peut supposer que ψ est anisotrope. Soient $B \in I^n F$ telle que $B + I^{n+1} F \in \overline{I^n}(F(\psi)/F)$, et $\varphi = B \otimes [1, t^{-1}]$. Alors, $\varphi \in I_q^{n+1} F(t)$ et $\varphi + I_q^{n+2} F \in \overline{I_q^{n+1}}(F(t)(\psi)/F(t))$. Par le corollaire 4.11, $\varphi \in I_q^{n+2} F(t)$. Le lemme 4.12 implique que $B \in I^{n+1} F$. $\square$

On obtient un corollaire qui va permettre de simplifier certains calculs plus tard :

Corollaire 4.14. *Soient $\varphi_0, \dots, \varphi_k$ des formes quadratiques totalement singulières anisotropes de dimension $> 2^n$ telles que φ_i soit anisotrope sur $F(\varphi_0)\cdots(\varphi_{i-1})$ pour tout $i \geq 1$. Soit $L = F(\varphi_0)\cdots(\varphi_k)$. Alors*:
(1) $\overline{I^j}(L/F) = \{0\}$ *pour tout* $j \leq n$.
(2) *Une forme bilinéaire métabolique sur L appartient à $I^{n+1}F$.*

Démonstration. Puisque $\dim \varphi_i > 2^j$ pour tous $j \leq n$ et $i \in \{0, \cdots, k\}$, on peut appliquer de manière répétée la proposition 4.13. $\square$

Démonstration du théorème 4.5. Soit $n \geq 1$ un entier.

(1) $I^nF \subset J_n^b(F)$: Soient $B \in I^nF$ non nulle, t une variable sur F et $\varphi = B \otimes [1, t^{-1}]$. Par le lemme 4.7, $\deg(\varphi) \leq \deg(B)+1$. Comme $\varphi \in I_q^{n+1}K$, on déduit par le théorème 4.10 que $\deg(\varphi) \geq n+1$, i.e., $\deg(B) \geq n$ et donc $B \in J_n^b(F)$.

(2) $J_n^b(F) \subset I^nF$: Soit $B \in J_n^b(F)$ non nulle. On peut supposer B anisotrope et on procède par induction sur $\dim B$. On a $\dim B \geq 2^n$.

Si $\dim B = 2^n$, alors $B_{F(B)}$ est métabolique. Par le théorème 4.1, B est semblable à une n-forme bilinéaire de Pfister, et donc $B \in I^nF$.

Si $\dim B > 2^n$. Puisque $\deg(B) = \deg(B_{F(B)})$, on obtient par induction que $B_{F(B)} \in I^nF(B)$. Par la proposition 4.13, on déduit que si $B \in I^kF$ avec $k < n$, alors $B \in I^{k+1}F$. Ainsi, en appliquant ceci de manière successive pour $k \in \{1, \dots, n-1\}$, on déduit que $B \in I^nF$. $\square$

Maintenant on donne un corollaire qui complète le lemme 4.7 :

Corollaire 4.15. *Soient t une variable sur F et B une forme bilinéaire non nulle. Alors*, $\deg(B) + 1 = \deg(B \otimes [1, t^{-1}])$.

Démonstration. Posons $\varphi = B \otimes [1, t^{-1}]$, $d = \deg(B)$ et $d' = \deg(\varphi)$. Si $d = 0$, alors il est clair que $d' = 1$. Supposons $d \geq 1$. L'inégalité $d' \leq d+1$ provient du lemme 4.7. Si $d+1 > d'$, alors $B \in J_{d'}^b(F) = I^{d'}F$. Par conséquent, $\varphi \in I_q^{d'+1}F$, une contradiction. $\square$

5. Les formes bilinéaires bonnes

5A. ***Généralités sur les formes bilinéaires bonnes.*** La théorie de déploiement met en évidence une classe importante de formes quadratiques, celle des formes quadratiques bonnes (“good forms” dans la terminologie de [Fitzgerald 1984]). Dans cette section, on va étendre certains résultats connus sur ces formes quadratiques au cas des formes bilinéaires en caractéristique 2.

Définition 5.1. Une forme bilinéaire non nulle est dite bonne si sa forme dominante est définie sur F.

Exemple 5.2. Si B est une forme bilinéaire de dimension paire et de déterminant $d \neq 1$, alors elle est bonne de degré 1 et de forme dominante $\langle 1, d\rangle_L$, où L est le corps dominant de B.

Démonstration. Soit π la forme dominante de B. Puisque $B_L \sim \pi$, les formes B_L et π ont le même déterminant (modulo un carré). Ainsi, π ne peut être qu'une 1-forme bilinéaire de Pfister et donc $\pi \simeq \langle 1, d\rangle_L$. □

Voici quelques propriétés générales liées aux formes bilinéaires bonnes :

Proposition 5.3. *Soit B une forme bilinéaire non nulle de hauteur h. On note $(F_i, B_i)_{0\leq i\leq h}$ la tour de déploiement standard de B et τ sa forme dominante. Posons* $\dim \tau = 2^d$.
(1) *Si B est bonne, alors il existe une unique d-forme bilinéaire de Pfister C définie sur F telle que $\tau \simeq C_{F_{h-1}}$.*
(2) *Si B est de dimension paire, alors B est bonne si et seulement si $B \perp C \in I^{d+1}F$ pour une certaine d-forme bilinéaire de Pfister C. Dans ce cas, $C_{F_{h-1}}$ est la forme dominante de B.*
(3) *Si B est de dimension impaire, alors B est bonne si et seulement si $B \perp \langle \det B\rangle$ est bonne. Dans ce cas, B et $B \perp \langle \det B\rangle$ ont la même forme dominante.*
(4) *Si* $\dim B$ *est paire et C est une d-forme bilinéaire de Pfister telle que $B \perp rC \in I^{d+2}F$ pour $r \in F^*$ convenable, alors $B_{h-1} \simeq (rC)_{F_{h-1}}$.*

Avant de prouver cette proposition, on donne un lemme préliminaire :

Lemme 5.4. *Soient B et C des formes bilinéaires anisotropes avec C une forme bilinéaire de Pfister. Soit t une variable sur F.*
(1) *Si $B \otimes [1, t^{-1}]$ est hyperbolique sur le corps de fonctions de $C \otimes [1, t^{-1}]$, alors il existe $\alpha_1, \dots, \alpha_s \in F^*$ des scalaires convenables tels que $B \simeq \alpha_1 C \perp \cdots \perp \alpha_s C$.*
(2) *Si* $\dim B = \dim C = 2^d$ *et $aB \perp C \in I^{d+1}F$ pour un certain $a \in F^*$, alors B est semblable à C.*

Démonstration. Posons $\varphi = B \otimes [1, t^{-1}]$ et $\psi = C \otimes [1, t^{-1}]$.

(1) On procède par induction sur $\dim B$. Soit $\alpha_1 \in D_F(B)D_F(C)$. Par le théorème 4.9, on a $\alpha_1\psi \subset \varphi$. Du lemme 4.6 on obtient $\alpha_1 C \subset B$. Soit B' une forme bilinéaire telle que $B \simeq \alpha_1 C \perp B'$. Puisque $(\varphi \perp \alpha_1\psi)_{\mathrm{an}} \simeq B' \otimes [1, t^{-1}]$ est hyperbolique sur $F(t)(\psi)$ et que $\dim B' < \dim B$, on obtient par induction que $B' \simeq \alpha_2 C \perp \cdots \perp \alpha_s C$ pour certains $\alpha_2, \dots, \alpha_s \in F^*$. Ainsi, $B \simeq \alpha_1 C \perp \cdots \perp \alpha_s C$.

(2) Puisque $aB \perp C \in I^{d+1}F$, on obtient que $a\varphi \perp \psi \in I^{d+2}F(t)$. Par le lemme 4.8, la forme φ devient hyperbolique sur le corps de fonctions de ψ. Par l'assertion (1), on déduit que B est semblable à C. □

Démonstration de la proposition 5.3. Soit B une forme bilinéaire non nulle de hauteur h, de degré d et de tour de déploiement standard $(F_i, B_i)_{0\leq i\leq h}$. Soit τ sa forme dominante.

La proposition est clairement vérifiée lorsque $h = 1$. On peut donc supposer $h \geq 2$.

(1) Supposons que B soit bonne. Soit ρ une forme bilinéaire définie sur F telle que $\rho_{F_{h-1}} \simeq \tau$. Alors $\rho_{F_{h-1}} \in I^d F_{h-1}$. Soit k le plus grand entier tel que $\rho \in I^k F$. Par le lemme 4.8, on a $k \leq d$. Si $k < d$, alors $\rho_{F_{h-1}} \in \overline{I^k}(F_{h-1}/F)$. Par le corollaire 4.14(1), $\rho \in I^{k+1}F$ car F_{h-1} est la succession de corps de fonctions de formes bilinéaires de dimension $> 2^d$, une contradiction. Ainsi, $\rho \in I^d F$ et par le lemme 4.8 ρ est semblable à une d-forme bilinéaire de Pfister C.

Il reste à prouver l'unicité de C. En effet, si C' est une autre forme bilinéaire de Pfister telle que $\tau \simeq C'_{F_{h-1}}$, alors $(C \perp C')_{F_{h-1}} \sim 0$. Le corollaire 4.14(2) implique que $C \perp C' \in I^{d+1}F$. Comme $C \perp C'$ est isotrope (car $1 \in D_F(C) \cap D_F(C')$), on obtient par le lemme 4.8 que $C \simeq C'$.

(2) Supposons que $\dim B$ soit paire. Soit C une d-forme bilinéaire de Pfister définie sur F.

Si $\tau \simeq C_{F_{h-1}}$, alors $B_{F_{h-1}} \perp C_{F_{h-1}} \in I^{d+1}F$. Puisque $B \perp C \in I^d F$, on déduit que $B \perp C + I^{d+1}F \in \overline{I^d}(F_{h-1}/F)$. Par le corollaire 4.14(1), on obtient $B \perp C \in I^{d+1}F$.

Réciproquement, si $B \perp C \in I^{d+1}F$, alors $x\tau \perp C_{F_{h-1}} \in I^{d+1}F_{h-1}$ pour $x \in F^*_{h-1}$ convenable. Par le lemme 5.4(2) et la multiplicativité d'une forme de Pfister, on déduit que τ est isométrique à $C_{F_{h-1}}$, et donc B est bonne.

(3) Supposons que $\dim B$ soit impaire. Posons $\alpha = \det B$.

Si B est bonne, alors par l'assertion (1) il existe C une d-forme bilinéaire de Pfister tel que $\tau \simeq C_{F_{h-1}}$. Comme B_{h-1} est semblable à la partie pure τ' de τ, on obtient par comparaison des déterminants que $\alpha B_{h-1} \simeq \tau'$. Ainsi, $\alpha B_{h-1} \perp \langle 1 \rangle \simeq C_{F_{h-1}}$, et donc $(\alpha B \perp \langle 1 \rangle \perp C)_{F_{h-1}} \sim 0$. Comme F_{h-1} est une succession de corps de fonctions de formes bilinéaires de dimension $\geq 2^d + 1$, on obtient par le corollaire 4.14(2) que $\alpha B \perp \langle 1 \rangle \perp C \in I^{d+1}F$. Ainsi, $\alpha B \perp \langle 1 \rangle$ est bonne, c'est-à-dire, $B \perp \langle \alpha \rangle$ est aussi bonne.

Réciproquement, supposons que $B \perp \langle \alpha \rangle$ soit bonne. Alors, il existe C une d-forme bilinéaire de Pfister tel que $B \perp \langle \alpha \rangle \perp C \in I^{d+1}F$. En particulier, $B_{h-1} \perp (\langle \alpha \rangle \perp C)_{F_{h-1}} \in I^{d+1}F_{h-1}$. Comme $B_{h-1} \perp \langle \alpha \rangle$ est semblable à τ, on obtient par le lemme 5.4 que $\tau \simeq C_{F_{h-1}}$ et donc B est bonne.

(4) Supposons que $B \perp rC \in I^{d+2}F$ pour un certain $r \in F^*$. Alors, $B_{h-1} \perp (rC)_{F_{h-1}} \in I^{d+2}F_{h-1}$. Par le lemme 4.8 on a que $B_{h-1} \simeq (rC)_{F_{h-1}}$. □

L'exemple suivant montre que la réciproque de l'assertion (4) de la proposition 5.3 n'est pas vraie en général :

Exemple 5.5. Soient F_0 un corps de caractéristique 2 et x, y des variable sur F_0. Soient $F = F_0(x, y)$ et $B = \langle 1, x, x + y, xy \rangle$. Alors :

(1) B est bonne de hauteur 2, de degré 1 et de forme dominante $\tau = \langle 1, y(x + y) \rangle$.

(2) On a $(B_{F(B)})_{\mathrm{an}} \simeq (y\tau)_{F(B)}$ mais $B \perp y\tau$ ne peut être dans I^3F.

Démonstration. Soit $\pi = \langle\langle x, y \rangle\rangle$. Puisque $\pi \in \mathscr{A}(\widetilde{B})$, la forme B est voisine de π. De plus, B est de déterminant $y(x + y) \neq 1$, donc B est nécessairement de hauteur

2, de degré 1 et de forme dominante τ. Puisque $\pi_{F(B)}$ est isotrope et $B \sim \pi \perp y\tau$, on obtient $(B_{F(B)})_{\mathrm{an}} \simeq (y\tau)_{F(B)}$. Mais par le lemme 4.8, on ne peut avoir $B \perp \alpha\tau \in I^3 F$ car B est anisotrope. □

5B. ***Formes bilinéaires bonnes de hauteur* 2.** La démonstration du lemme suivant se fait comme dans le cas des formes quadratiques en utilisant la multiplicativité d'une forme bilinéaire de Pfister :

Lemme 5.6. *Soient $\alpha_1, \dots, \alpha_s \in F^*$ et B une forme bilinéaire de Pfister. Alors, $i_W(\perp_{i=1}^{s} \alpha_i B) = 0$ ou $\geq \dim B$.*

On étend aux formes bilinéaires un résultat récent de Karpenko [2004] décrivant les dimensions de certaines formes quadratiques de I^n sur un corps de caractéristique $\neq 2$:

Proposition 5.7. *Soient $n \geq 1$ un entier et $B \in I^n F$ anisotrope telle que* $\dim B < 2^{n+1}$. *Alors*, $\dim B \in \{2^{n+1} - 2^i \mid 1 \leq i \leq n+1\}$.

Démonstration. Soit t une variable sur F. On a $B \otimes [1, t^{-1}] \in I_q^{n+1} F(t)$ qui est de dimension $< 2^{n+2}$. En considérant $F(t)$ comme étant le corps résiduel d'un corps K de caractéristique 0, complet pour une valuation discrète, on peut utiliser le résultat de Karpenko [2004] pour déduire qu'un relèvement de $B \otimes [1, t^{-1}]$ à K a pour dimension $2^{n+2} - 2^i$ avec $1 \leq i \leq n+2$ (en fait on a ceci pour $2 \leq i \leq n+2$ puisque $\dim B$ est paire). Ainsi, $\dim B \in \{2^{n+1} - 2^i \mid 1 \leq i \leq n+1\}$. □

Remarque 5.8. La même idée de la preuve de la proposition 5.7 permet d'avoir l'analogue du résultat de Karpenko dans le cas des formes non singulières. Mais on n'en a pas besoin ici.

Maintenant on donne nos résultats classifiant les formes bilinéaires bonnes de hauteur 2. Dans le cas d'une forme de dimension impaire on obtient :

Proposition 5.9. *Soit B une forme bilinéaire anisotrope bonne de hauteur* 2 *et de dimension impaire. Alors*:
(1) *B est une voisine de Pfister.*
(2) *Il existe π semblable à une forme bilinéaire de Pfister dont B est voisine*, *et C une forme bilinéaire semblable à la partie pure d'une certaine d-forme bilinéaire de Pfister telles que*: $\dim B > 2^d$ *et $B \sim C \perp \pi$.*
Réciproquement, *une forme bilinéaire B vérifiant les conditions de* (2) *est bonne de hauteur* 2.

Démonstration. Soit B une forme bilinéaire anisotrope de dimension impaire et de hauteur 2. Soit $(B_i, F_i)_{0 \leq i \leq 2}$ sa tour de déploiement standard. Par le corollaire 4.2, il existe une unique forme bilinéaire de Pfister τ définie sur F tel que $B_1 \simeq \alpha\tau'$ pour un certain $\alpha \in F_1^*$. En comparant les déterminants, on peut supposer que $\alpha \in F^*$. De plus, si $\dim \tau = 2^d$ alors $\dim B \geq 2^d + 1 > 2^d$ puisque $\dim B > \dim B_1 = 2^d - 1$.

Posons $C = \alpha\tau'$. Puisque $(B \perp C)_{F(B)} \sim 0$ et $\dim B > \dim C$, on obtient par la proposition 3.9 que la forme $(B \perp C)_{\mathrm{an}}$ est semblable à une forme bilinéaire de Pfister dont B est voisine. La forme π sera alors $(B \perp C)_{\mathrm{an}}$. D'où le résultat.

Réciproquement, si B est une forme bilinéaire qui vérifie les conditions de (2), alors $B_{F(B)} \sim C_{F(B)}$. Puisque $\dim B > 2^d > \dim C$, on obtient par [Hoffmann and Laghribi 2006] que la forme $C_{F(B)}$ est anisotrope. Ainsi, $(B_{F(B)})_{\mathrm{an}} \simeq C_{F(B)}$. Comme C est semblable à la partie pure d'une forme bilinéaire de Pfister, on a que $C_{F(B)}$ est de hauteur 1, et donc B est bonne de hauteur 2. □

Le résultat suivant classifie les formes bilinéaires bonnes de hauteur 2 et de dimension paire :

Théorème 5.10. *Soit B une forme bilinéaire non nulle de dimension paire, de hauteur 2 et de degré d. On suppose que B est bonne de forme dominante C.*
(1) *Si* $\dim B > 2^{d+1}$, *alors on a deux cas*:
(i) *soit* $\dim B$ *est une puissance de* 2. *Dans ce cas, on ne peut pas conclure.*
(ii) *soit* $\dim B$ *n'est pas une puissance de* 2. *Dans ce cas, il existe ρ une forme bilinéaire de dimension impaire tel que $B \simeq \rho \otimes C$ et $B \perp \alpha C$ soit semblable à une n-forme bilinéaire de Pfister avec $n \geq d+2$ et $\alpha = \det \rho$.*
(2) *Si* $\dim B \leq 2^{d+1}$, *alors* $\dim B = 2^{d+1}$ *et $B \sim xC \perp \pi$ avec $x \in F^*$ et π une forme semblable à une $(d+1)$-forme bilinéaire de Pfister.*
Réciproquement, les conditions dans (1)(ii) (*resp. dans* (2)) *sont suffisantes pour dire que B est bonne de hauteur* 2, *de degré d et de forme dominante C.*

Démonstration. Soient t une variable sur F, $L = F(t)$, $\varphi = B \otimes [1, t^{-1}]$ et $\tau = C \otimes [1, t^{-1}]$. Puisque $(B_{F(B)})_{\mathrm{an}} \simeq a(C_{F(B)})$ pour un certain $a \in F(B)^*$, alors $\varphi_{L(B)} \sim a(\tau_{L(B)})$. Ainsi,

$$\varphi_{L(\tau)(B)} \sim 0. \tag{5}$$

En particulier, $\varphi_{L(\tau)(\varphi)(B)} \sim 0$. Ceci implique que

$$\varphi_{L(\tau)(\varphi)} \sim 0 \tag{6}$$

car sinon la forme $\widetilde{B}_{L(\tau)(\varphi)}$ (qui est anisotrope par [Laghribi 2002a]) serait dominée par $(\varphi_{L(\tau)(\varphi)})_{\mathrm{an}}$, ce qui n'est pas possible par le théorème 4.9 car $\dim(\varphi_{L(\tau)(\varphi)})_{\mathrm{an}} < \dim \varphi = 2 \dim \widetilde{B}$.

(1) Supposons que $\dim B > 2^{d+1}$ et que $\dim B$ ne soit pas une puissance de 2. En particulier, $\dim \varphi$ n'est pas une puissance de 2. La forme $\varphi_{L(\tau)}$ est nécessairement isotrope et donc par (6) $\varphi_{L(\tau)} \sim 0$, car sinon $\varphi_{L(\tau)}$ serait de hauteur 1 et donc serait semblable à une forme de Pfister [Laghribi 2002b], en particulier $\dim \varphi$ serait une puissance de 2. Par le lemme 5.4 on obtient $B \simeq \rho \otimes C$ pour une certaine forme bilinéaire ρ. On a $\dim \rho$ impaire car sinon B serait dans $I^{d+1}F$, ce qui contredirait $\deg(B) = d$. Pour $\alpha = \det \rho$, on a $\rho \perp \langle \alpha \rangle \in I^2 F$, ainsi $B \perp \alpha C \in I^{d+2} F$. Puisque

$(B_{F(B)})_{\mathrm{an}}$ est semblable à $C_{F(B)}$, on obtient par le lemme 4.8 que $(B \perp \alpha C)_{F(B)} \sim 0$. La forme $B \perp \alpha C$ est anisotrope, car sinon par le lemme 5.6 $i_W(B \perp \alpha C) \geq \dim C = 2^d$, et donc on aurait $\dim(B \perp \alpha C)_{\mathrm{an}} \leq \dim B - \dim C < \dim B$. Le théorème 3.5 impliquerait $B \perp \alpha C \sim 0$ et donc B serait isotrope, une contradiction. Par la proposition 3.9, $B \perp \alpha C$ est semblable à une n-forme de Pfister. Puisque $\dim(B \perp \alpha C) > 2 \dim C = 2^{d+1}$, on obtient $n \geq d+2$.

(2) Supposons que $\dim B \leq 2^{d+1}$.

(i) Cas où $B_{F(C)}$ est anisotrope. Puisque $B_{F(B)(C)}$ est métabolique, on déduit par le théorème 4.1 que $B_{F(C)}$ est semblable à une forme bilinéaire de Pfister. Ainsi, $\dim B = 2^{d+1}$ puisque $\dim B > 2^d$.

(ii) Cas où $B_{F(C)}$ est isotrope. Par la relation (5) et [Laghribi 2005, prop. 3.9], on obtient que $\varphi_{L(\tau)(C)}$ est hyperbolique. Comme $\tau_{L(C)}$ est isotrope, l'extension $L(C)(\tau)/L(C)$ est transcendante pure et donc $\varphi_{L(C)}$ est hyperbolique. Par le lemme 4.6, on a que $B_{F(C)}$ est métabolique. Puisque $2^d < \dim B \leq 2^{d+1}$, on déduit de [Laghribi 2005, th. 1.2] que $B \simeq \beta_1 B_1 \perp \beta_2 B_2$, où B_1, B_2 sont des d-formes bilinéaires de Pfister associées à $\widetilde{C}$. En particulier, $\dim B = 2^{d+1}$.

Ainsi, dans les deux cas (i) et (ii) on a que $\dim B = 2^{d+1}$.

Soit $x \in D_F(B)$. On a $0 < \dim(B \perp xC)_{\mathrm{an}} < 2^{d+1} + 2^d$ du fait que $B \perp xC$ est isotrope mais non métabolique. Puisque $B \perp xC \in I^{d+1}F$, on déduit par la proposition 5.7 que $\dim(B \perp xC)_{\mathrm{an}} = 2^{d+1}$. Par le lemme 4.8, il existe π semblable à une $(d+1)$-forme bilinéaire de Pfister tel que $B \perp xC \sim \pi$.

Réciproquement, si on est dans le cas (1)(ii) ou (2), alors $B \perp C \in I^{d+1}F$. La proposition 5.3 implique que B est bonne de degré d et de forme dominante C. Il reste à prouver que B est de hauteur 2. En effet, dans le cas (1)(ii) on a $B_{F(B)} \sim (\alpha C)_{F(B)}$. Puisque $\dim C = 2^d < \dim B$, la forme $C_{F(B)}$ est anisotrope [Hoffmann and Laghribi 2006]. Ainsi, $(B_{F(B)})_{\mathrm{an}} \simeq (\alpha C)_{F(B)}$ et donc B est de hauteur 2. Si on est dans le cas (2), alors $B_{F(B)(C)} \sim \pi_{F(B)(C)}$. Comme $\dim B = \dim \pi$ et $B_{F(B)}$ est isotrope, la forme $\pi_{F(B)(C)}$ est isotrope et donc métabolique. Ainsi, $B_{F(B)(C)} \sim 0$. La forme $B_{F(B)}$ ne peut être métabolique, car sinon B serait semblable à une $(d+1)$-forme de Pfister et donc $xC \in I^{d+1}F$, ce qui contredirait l'anisotropie de C. Ainsi, $0 < \dim(B_{F(B)})_{\mathrm{an}} < 2^{d+1}$. Puisque $B_{F(B)(C)} \sim 0$, on a par [Laghribi 2005, th. 1.2] que $(B_{F(B)})_{\mathrm{an}}$ est semblable à une d-forme bilinéaire de Pfister et donc B est de hauteur 2. □

On n'a pas une caractérisation des formes bilinéaires indiquées dans l'assertion (1)(i) du théorème 5.10, c'est-à-dire, les formes B anisotropes qui sont bonnes de hauteur 2 telles que $\dim B$ soit une puissance de 2 strictement supérieure à $2^{\deg(B)+1}$. Ci-dessous on donne quelques exemples de telles formes bilinéaires.

Exemple 5.11. Soient $k \geq 2$ et $d \geq 1$ des entiers et $x_1, \ldots, x_{d-1}, y_1, \ldots, y_{k-1}, u, v$ des variables sur un corps F_0 de caractéristique 2. Soit $F = F_0(u, v, x_i, y_j)$. On considère $C = \langle\langle x_1, \ldots, x_{d-1}\rangle\rangle$, $D = \langle\langle y_1, \ldots, y_{k-1}\rangle\rangle$ et $\pi = \langle\langle u, v\rangle\rangle \otimes D \otimes C$. Alors, la forme

$$B = \langle 1, u, u+v, uv\rangle \otimes C \perp \langle\langle u, v\rangle\rangle \otimes D' \otimes C$$

vérifie les conditions suivantes :

(1) B est anisotrope de dimension 2^{k+d}.
(2) B est une voisine de π.
(3) $B \sim \pi \perp v\,\langle 1, v(u+v)\rangle \otimes C$.
(4) B est bonne de forme dominante $\langle 1, v(u+v)\rangle \otimes C$, de hauteur 2 et de degré d.

Démonstration. La forme bilinéaire $\langle 1, u, u+v, uv\rangle$ est associée à la forme quadratique totalement singulière $[1] \perp [u] \perp [v] \perp [uv]$. Ainsi, on déduit facilement que π est associée à $\widetilde{B}$, et par conséquent B est une voisine de π. Puisque π est anisotrope, la forme B est aussi anisotrope. Aussi, on vérifie que

$$B \sim \pi \perp v\,\langle 1, v(u+v)\rangle \otimes C.$$

Par raison de dimension et le théorème 3.5, la forme $\langle 1, v(u+v)\rangle \otimes C$ est anisotrope sur $F(B)$. Puisque $\pi_{F(B)}$ est isotrope, on déduit que

$$(B_{F(B)})_{\mathrm{an}} \simeq (v\,\langle 1, v(u+v)\rangle \otimes C)_{F(B)}.$$

Ainsi, B est bonne de hauteur 2, de degré d et de forme dominante

$$\langle 1, v(u+v)\rangle \otimes C. \qquad \square$$

Question 5.12. Soit B une forme bilinéaire anisotrope bonne de hauteur 2, de degré d et de forme dominante C. On suppose que $\dim B = 2^k > 2^{d+1}$. A-t-on B voisine d'une k-forme bilinéaire de Pfister π et $B \sim \alpha\pi \perp \beta C$ pour certains scalaires $\alpha, \beta \in F^*$?

5C. ***Sur les dimensions des formes bilinéaires de hauteur 2.*** Avant de donner notre résultat concernant les dimensions des formes bilinéaires de hauteur 2, on commence par donner un résultat sur la décomposition de Witt des formes bilinéaires. Notons tout d'abord que la forme métabolique M intervenant dans la décomposition (2) n'est pas unique à isométrie près. En voici un exemple :

Exemple 5.13. Soit t une variable sur F. On a

$$\langle t:1:0\rangle \perp \langle t\rangle \simeq \langle 0:1:0\rangle \perp \langle t\rangle$$

et $\langle t:1:0\rangle \not\simeq \langle 0:1:0\rangle$.

Démonstration. Posons $B = \langle t:1:0\rangle \perp \langle t\rangle$. Soit $\{e, f, g\}$ une $F(t)$-base de l'espace sous-jacent à B telle que :

$$\begin{cases} B(e,e) = B(g,g) = t, \\ B(f,f) = B(e,g) = B(f,g) = 0, \\ B(e,f) = 1. \end{cases}$$

La restriction de B à l'espace engendré par $\{e+g,\ f\}$ est la forme $\langle 0:1:0\rangle$. Ainsi, $\langle t:1:0\rangle \perp \langle t\rangle \simeq \langle 0:1:0\rangle \perp \langle \alpha\rangle$ pour un certain $\alpha \in F(t)^*$. En comparant les déterminants dans cette dernière isométrie, on trouve que $a = \alpha$ (modulo un carré). Finalement, $\langle t:1:0\rangle \not\simeq \langle 0:1:0\rangle$ puisque les formes quadratiques associées à ces formes bilinéaires sont $[t] \perp [0]$ et $[0] \perp [0]$ qui ne sont pas isométriques. $\square$

Définition 5.14. Une forme bilinéaire est dite un plan hyperbolique si elle est isométrique à $\langle 0:1:0\rangle$. On la note $\mathcal{H}$.

La proposition suivante raffine la décomposition de Witt d'une forme bilinéaire.

Proposition 5.15. *Soit B une forme bilinéaire de dimension ≥ 1. Alors, il existe un couple d'entiers (s,t) unique, des scalaires $a_1,\dots,a_t \in F^*$ tels que:*
(1) $B \simeq s \times \mathcal{H} \perp (\langle a_1:1:0\rangle \perp \cdots \perp \langle a_t:1:0\rangle) \perp B_{an}$.
(2) La forme quadratique $(\widetilde{B})_{an}$ est associée à la forme bilinéaire $\langle a_1,\dots,a_t\rangle \perp B_{an}$.
En particulier, $\langle a_1,\dots,a_t\rangle \perp B_{an}$ est anisotrope, $t + \dim B_{an} = \dim(\widetilde{B})_{an}$ et $\dim \widetilde{B} - \dim(\widetilde{B})_{an} = 2s+t$.

Démonstration. Soit M une forme métabolique telle que $B \simeq M \perp B_{\mathrm{an}}$. Parmi toutes ces décompositions, on choisit celle telle que M contienne comme sous-forme un nombre maximal de copies de $\mathcal{H}$. Notons ce nombre s. Alors,

$$B \simeq s \times \mathcal{H} \perp (\langle a_1:1:0\rangle \perp \cdots \perp \langle a_t:1:0\rangle) \perp B_{\mathrm{an}} \tag{7}$$

pour certains $a_1,\dots,a_t \in F^*$ avec $\langle a_1,\dots,a_t\rangle$ anisotrope. Sinon, il existerait $l_1,\dots,l_t \in F$ non tous nuls tels que $b := \sum_{i=1}^t a_i l_i^2 \in D_F(B_{\mathrm{an}})$. Ainsi, $B_{\mathrm{an}} \simeq \langle b\rangle \perp C$ pour une certaine forme bilinéaire C, et on vérifie facilement que

$$\langle a_1:1:0\rangle \perp \cdots \perp \langle a_t:1:0\rangle \simeq \langle b:1:0\rangle \perp M' \tag{8}$$

pour une forme bilinéaire convenable M'. Par unicité de la partie anisotrope, la forme M' est nécessairement métabolique. En utilisant la même preuve que celle de l'exemple 5.13, on a

$$\langle b:1:0\rangle \perp \langle b\rangle \simeq \mathcal{H} \perp \langle b\rangle \tag{9}$$

En substituant (8) et (9) dans (7), on aurait une contradiction avec le choix de s.

Posons $C = \langle a_1,\dots,a_t\rangle \perp B_{\mathrm{an}}$. En prenant la forme quadratique associée à B, on obtient l'isométrie :

$$\widetilde{B} \simeq (2s+t) \times [0] \perp \widetilde{C} \simeq j \times [0] \perp (\widetilde{B})_{\mathrm{an}}$$

pour un certain $j \geq 0$. Puisque $\widetilde{C}$ est anisotrope, on obtient par la simplification de Witt [Hoffmann and Laghribi 2004, lem. 2.6] que $\widetilde{C} \simeq (\widetilde{B})_{\mathrm{an}}$. En particulier, $j = 2s + t = \dim B - \dim(\widetilde{B})_{\mathrm{an}}$. D'où l'unicité du couple (s, t). □

Notations 5.16. Comme dans la proposition 5.15, on note $i_h(B) = s$ et $i_m(B) = t$.

Comme prouvé par Milnor [1971, th. 3], une forme bilinéaire B est déterminée, à isométrie près, par sa dimension qui vaut $2\dim(\widetilde{B})_{\mathrm{an}} - \dim B_{\mathrm{an}} + 2i_h(B)$, sa partie anisotrope et le F^2-espace vectoriel $D_F(B) \cup \{0\}$.

La proposition 5.15 se précise beaucoup plus dans le cas d'une forme bilinéaire voisine de Pfister étendue à son corps de fonctions :

Proposition 5.17. *Soit B une forme bilinéaire anisotrope qui est une voisine de Pfister. Posons* $\dim B = 2^n + l$ *avec* $0 < l \leq 2^n$, $B_1 = (B_{F(B)})_{an}$, $s = i_h(B_{F(B)})$ *et* $t = i_m(B_{F(B)})$. *Alors, on a*:
(1) $2s + t = l$ *et* $t + \dim B_1 = 2^n$.
(2) $\dim B = 2^{n+1} - \dim B_1 + 2s$.

Démonstration. Posons $C = (B_{F(B)})_{\mathrm{an}}$. Par la proposition 5.15, on peut écrire :

$$B_{F(B)} \simeq s \times \mathcal{H} \perp (\langle a_1 : 1 : 0\rangle \perp \cdots \perp \langle a_t : 1 : 0\rangle) \perp C \tag{10}$$

avec $\langle a_1, \ldots, a_t\rangle \perp C$ anisotrope. En considérant la forme quadratique associée à $B_{F(B)}$, on obtient :

$$\widetilde{B}_{F(B)} \simeq (2s + t) \times [0] \perp ([a_1] \perp \cdots \perp [a_t]) \perp \widetilde{C}. \tag{11}$$

Puisque B est une voisine de Pfister, la forme $\widetilde{B}$ est une quasi-voisine de Pfister. D'après [Laghribi 2004a, th. 3.6] et [Hoffmann and Laghribi 2004, section 8], on sait que

$$\widetilde{B}_{F(B)} \simeq l \times [0] \perp (\widetilde{B}_{F(B)})_{\mathrm{an}} \tag{12}$$

et donc $\dim(\widetilde{B}_{F(B)})_{\mathrm{an}} = 2^n$. Par la proposition 5.15, on a $t + \dim B_1 = 2^n$ et $l = 2s + t$, d'où l'affirmation (1). (2) se déduit de (1). □

De cette proposition on peut déduire quelques corollaires. Le premier donne des exemples de formes bilinéaires dont la décomposition de Witt admet des plans métaboliques non hyperboliques et inversement :

Corollaire 5.18. (1) *Si B est une n-forme bilinéaire de Pfister anisotrope, alors* $i_h(B_{F(B)}) = 0$ *et* $i_m(B_{F(B)}) = 2^{n-1}$.
(2) *Si B est une forme bilinéaire de dimension* 6 *comme dans l'exemple 3.12, alors* $i_h(B_{F(B)}) = 1$ *et* $i_m(B_{F(B)}) = 0$.

Démonstration. Puisque dans les deux cas les formes bilinéaires sont des voisines de Pfister, on applique la proposition 5.17 sachant que dans le premier cas on a $\dim(B_{F(B)})_{\mathrm{an}} = 0$, et dans le deuxième cas on a $\dim(B_{F(B)})_{\mathrm{an}} = 4$. □

Corollaire 5.19. *Soient B une forme bilinéaire anisotrope et K comme dans la proposition 3.13. Posons* $\dim B = 2^n + l$ *avec* $0 < l \leq 2^n$, $s = i_h(B_{K(B)})$ *et* $t = i_m(B_{K(B)})$. *Alors, on a*:
(1) $2s + t = l$ *et* $t + \dim(B_{K(B)})_{an} = 2^n$.
(2) $\dim B = 2^{n+1} - \dim(B_{K(B)})_{an} + 2s$.

Démonstration. Puisque B_K est une voisine bilinéaire de Pfister anisotrope, le corollaire se déduit de la proposition 5.17. □

Voici notre résultat autour des dimensions des formes bilinéaires anisotropes de hauteur 2 :

Corollaire 5.20. *Soit B une forme bilinéaire anisotrope de hauteur 2 et de forme dominante de dimension 2^d, non nécessairement bonne. Supposons que $2^n <$* $\dim B$ $\leq 2^{n+1}$ *pour un certain entier $n \geq 1$. Alors,* $\dim B$ *appartient à l'ensemble* $\{2^{n+1} - 2^d + 2s + \epsilon \mid 0 \leq s \leq 2^{d-1} - \epsilon\}$, *où $\epsilon = 0$ ou 1 suivant que* $\dim B$ *est paire ou impaire.*

Démonstration. Soit ϵ comme dans le corollaire, et K comme dans la proposition 3.13. Si $\dim B = 2^{n+1} - \epsilon$, alors dans ce cas $\dim B \in \{2^{n+1} - 2^d + 2s + \epsilon \mid 0 \leq s \leq 2^{d-1} - \epsilon\}$ pour $s = 2^{d-1} - \epsilon$. On peut donc supposer que $\dim B \neq 2^{n+1} - \epsilon$. Posons $C = (B_{K(B)})_{\mathrm{an}}$ et $B_1 = (B_{F(B)})_{\mathrm{an}}$. Alors, $B_{K(B)} \sim (B_1)_{K(B)}$. La forme $(B_1)_{K(B)}$ est anisotrope, car sinon $\dim((B_1)_{K(B)})_{\mathrm{an}} = \epsilon$ et donc B_K serait de hauteur 1, en particulier on aurait $\dim B = 2^{n+1} - \epsilon$. Ainsi. $C \simeq (B_1)_{K(B)}$ et donc $\dim C = 2^d - \epsilon$. Par le corollaire 5.19(2), on a $\dim B = 2^{n+1} - 2^d + 2s + \epsilon$ où $s = i_h(B_{K(B)})$. Puisque $\dim B < 2^{n+1} - \epsilon$, on déduit que $0 \leq s < 2^{d-1} - \epsilon$. D'où le corollaire. □

Remarque 5.21. Dans le cas des formes bilinéaires B anisotropes de hauteur 2 et de degré $d > 0$, le corollaire 5.20 prend son intérêt lorsque $\dim B \geq 2^{d+1}$, puisque dans le cas $\dim B < 2^{d+1}$ la proposition 5.7 donne de manière précise la dimension de B.

On finit par un commentaire sur les dimensions des formes bilinéaires anisotropes de hauteur 2. On sait par le corollaire 5.20 qu'une telle forme de degré 2 ne peut avoir que la dimension 2^n, $2^n - 4$ ou $2^n - 2$. Le théorème 5.10 montre que les deux premiers entiers sont réalisables comme dimensions de formes bilinéaires anisotropes (bonnes) de hauteur et degré 2. En ce qui concerne l'entier $2^n - 2$, on a montré dans un travail en préparation [Laghribi and Rehmann $\geq$ 2007] qu'une forme de dimension $2^n - 2$ de hauteur et degré 2 qui n'est pas bonne a exactement la dimension 6 et que l'entier 8 est aussi réalisable comme dimension d'une forme bilinéaire de hauteur et degré 2. En ce qui concerne le degré $d > 2$, on ne sait pas exactement lesquelles des possibilités données par le corollaire 5.20 sont réalisables comme dimensions de formes bilinéaires anisotropes de hauteur 2 et de degré d.

Remerciements

Cet article a été achevé durant un séjour à l'Universität Bielefeld pour la période Mars-Mai 2006. Je tiens à remercier Ulf Rehmann d'avoir permis ce séjour et pour les discussions que nous avons eues autour du sujet de cet article. Je remercie aussi le Sonderforschungsbereich 701 "Spektrale Strukturen und topologische Methoden in der Mathematik" pour le soutien offert durant ce séjour.

Bibliographie

[Aravire and Baeza 2003] R. Aravire and R. Baeza, "The behavior of quadratic and differential forms under function field extensions in characteristic two", *J. Algebra* **259**:2 (2003), 361–414. MR 2003k:11060 Zbl 01891328

[Baeza 1973] R. Baeza, "Ein Teilformensatz für quadratische Formen in Charakteristik 2", *Math. Z.* **135** (1973), 175–184. MR 49 #2534 Zbl 0263.15015

[Baeza 1978] R. Baeza, *Quadratic forms over semilocal rings*, Lecture Notes in Mathematics **655**, Springer, Berlin, 1978. MR 58 #10972 Zbl 0382.10014

[Fitzgerald 1981] R. W. Fitzgerald, "Function fields of quadratic forms", *Math. Z.* **178**:1 (1981), 63–76. MR 83b:10021 Zbl 0442.10013

[Fitzgerald 1984] R. W. Fitzgerald, "Quadratic forms of height two", *Trans. Amer. Math. Soc.* **283**:1 (1984), 339–351. MR 85f:11027 Zbl 0512.10011

[Hoffmann and Laghribi 2004] D. W. Hoffmann and A. Laghribi, "Quadratic forms and Pfister neighbors in characteristic 2", *Trans. Amer. Math. Soc.* **356** (2004), 4019–4053. MR 2005e:11041 Zbl 02091232

[Hoffmann and Laghribi 2006] D. W. Hoffmann and A. Laghribi, "Isotropy of quadratic forms over the function field of a quadric in characteristic 2", *J. Algebra* **295**:2 (2006), 362–386. MR 2007a:11052 Zbl 05019786

[Karpenko 2004] N. A. Karpenko, "Holes in I^n", *Ann. Sci. École Norm. Sup.* (4) **37**:6 (2004), 973–1002. MR 2005k:11082 Zbl 02174960

[Knebusch 1970] M. Knebusch, "Grothendieck- und Wittringe von nichtausgearteten symmetrischen Bilinearformen", *S.-B. Heidelberger Akad. Wiss. Math.-Natur. Kl.* **1969/70** (1970), 93–157. MR 42 #6001 Zbl 0256.15016

[Knebusch 1976] M. Knebusch, "Generic splitting of quadratic forms. I", *Proc. London Math. Soc.* (3) **33**:1 (1976), 65–93. MR 54 #230 Zbl 0351.15016

[Knebusch 1977] M. Knebusch, "Generic splitting of quadratic forms. II", *Proc. London Math. Soc.* (3) **34**:1 (1977), 1–31. MR 55 #379 Zbl 0359.15013

[Knebusch $\geq$ 2007] M. Knebusch, "Spezialisierung von quadratischen und symmetrisch bilinearen Formen". livre en préparation.

[Knebusch and Rehmann 2000] M. Knebusch and U. Rehmann, "Generic splitting towers and generic splitting preparation of quadratic forms", pp. 173–199 in *Quadratic forms and their applications* (Dublin, 1999), edited by E. Bayer-Fluckiger et al., Contemp. Math. **272**, Amer. Math. Soc., Providence, RI, 2000. MR 2001g:11052 Zbl 0995.11028

[Laghribi 2002a] A. Laghribi, "Certaines formes quadratiques de dimension au plus 6 et corps des fonctions en caractéristique 2", *Israel J. Math.* **129** (2002), 317–361. MR 2003f:11047 Zbl 1049.11040

[Laghribi 2002b] A. Laghribi, "On the generic splitting of quadratic forms in characteristic 2", *Math. Z.* **240**:4 (2002), 711–730. MR 2003f:11048 Zbl 1007.11017

[Laghribi 2004a] A. Laghribi, "On splitting of totally singular quadratic forms", *Rend. Circ. Mat. Palermo* (2) **53**:3 (2004), 325–336. MR 2006f:11040 Zbl 1097.11017

[Laghribi 2004b] A. Laghribi, "Quasi-hyperbolicity of totally singular quadratic forms", pp. 237–248 in *Algebraic and arithmetic theory of quadratic forms*, Contemp. Math. **344**, Amer. Math. Soc., Providence, RI, 2004. MR 2005f:11066 Zbl 02154490

[Laghribi 2005] A. Laghribi, "Witt kernels of function field extensions in characteristic 2", *J. Pure Appl. Algebra* **199**:1-3 (2005), 167–182. MR 2006c:11041 Zbl 02175024

[Laghribi and Rehmann $\geq$ 2007] A. Laghribi and U. Rehmann, "Bilinear forms of height and degree 2 in characteristic 2". En préparation.

[Milnor 1971] J. Milnor, "Symmetric inner products in characteristic 2", pp. 59–75 in *Prospects in mathematics* (Princeton, 1970), edited by F. Hirzebruch et al., Ann. of Math. Studies **70**, Princeton Univ. Press, Princeton, N.J., 1971. MR 50 #367 Zbl 0265.10013

[Milnor and Husemoller 1973] J. Milnor and D. Husemoller, *Symmetric bilinear forms*, Ergebnisse der Mathematik **73**, Springer, New York, 1973. MR 58 #22129 Zbl 0292.10016

Received June 12, 2006. Revised October 9, 2006.

AHMED LAGHRIBI
FACULTÉ JEAN PERRIN
LABORATOIRE DE MATHÉMATIQUES DE LENS EA2462
FÉDÉRATION CNRS NORD-PAS-DE-CALAIS FR2956
RUE JEAN SOUVRAZ - SP18
F-62307 LENS
FRANCE
laghribi@euler.univ-artois.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 232, No. 1, 2007

ASYMPTOTICALLY LINEAR HAMILTONIAN SYSTEMS WITH LAGRANGIAN BOUNDARY CONDITIONS

CHUN-GEN LIU

We study the multiplicity of the solutions of certain asymptotically linear Hamiltonian systems with a Lagrangian boundary condition.

1. Introduction and main results

We consider the solutions of the nonlinear Hamiltonian systems with Lagrangian boundary condition

$$\dot{x}(t) = JH'(t, x(t)), \qquad x(0) \in L, \quad x(1) \in L. \tag{1–1}$$

where $x(t) \in \mathbb{R}^{2n}$ and

$$J = \begin{pmatrix} 0 & -I_n \\ I_n & 0 \end{pmatrix}$$

is the standard symplectic matrix with I_n the identity in $\mathbb{R}^n$, and $L \in \Lambda(n)$, where $\Lambda(n)$ is the set of all Lagrangian subspaces of $(\mathbb{R}^{2n}, \omega_0)$ with standard symplectic form $\omega_0 = \sum_{j=1}^n dx_j \wedge dy_j$. The Hamiltonian function $H \in C^2([0, 1] \times \mathbb{R}^{2n}, \mathbb{R})$ satisfies these conditions:

- (H_0): $H'(t, 0) \equiv 0$, $t \in [0, 1]$.
- (H_∞): There exist continuous symmetric matrix functions $B_1(t)$ and $B_2(t)$ with $i_L(B_1) = i_L(B_2)$, $\nu_L(B_2) = 0$ such that

  $$B_1(t) \le H''(t, x) \le B_2(t)$$

  for all (t, x) with $|x| \ge r$ for some large $r > 0$ and for all $t \in [0, 1]$.

For two symmetric matrices A and B, $A \ge B$ means that $A - B$ is a semipositive definite matrix, and $A > B$ similarly means that $A - B$ is a positive definite matrix.

For a Lagrangian subspace L of the standard symplectic vector space $(\mathbb{R}^{2n}, \omega_0)$, [Liu 2007] defined the Maslov-type index pair $(i_L(B), \nu_L(B)) \in \mathbb{Z} \times \{0, 1, \dots, n\}$ for a continuous symmetric matrix function $B : [0, 1] \to \mathrm{L}_s(2n)$ (here $\mathrm{L}_s(2n)$ is the

---

MSC2000: 58E05, 34C25.
Keywords: index theory, Hamiltonian systems, Lagrangian boundary conditions.
Partially supported by the National Natural Science Foundation of China, 973 program of MOST.

set of symmetric $2n \times 2n$ matrices). In the Appendix, we give a brief introduction of this index theory.

Theorem 1.1. *Let H satisfy conditions (H_0) and (H_∞). Suppose $JB_1(t) = B_1(t)J$ and $B_0(t) = H''(t, 0)$ satisfying one of the twisted conditions*

$$B_1(t) + kI \le B_0(t), \tag{1–2}$$

$$B_0(t) + kI \le B_1(t), \tag{1–3}$$

for some constant $k \ge \pi$. Then (1–1) *possesses at least one nontrivial solution. If $\nu_L(B_0) = 0$, the system* (1–1) *possesses at least two nontrivial solutions.*

For the periodic solutions of a asymptotically linear Hamiltonian system, we refer to [Chang 1981; Long 1993; Conley and Zehnder 1984; Liu 2005b]. We note that we only need to prove the case

$$L = L_0 = \{0\} \oplus \mathbb{R}^n.$$

The reason is that there is an orthogonal symplectic matrix P such that $PL = L_0$. All the conditions hold in (1–1) after taking $z(t) = Px(t)$ there. We note that the problem (1–1) is related to the Bolza problem (see [Clarke and Ekeland 1982; Ekeland 1990]).

We should briefly review the general study of the problem (1–1). For a general symplectic manifold (M, ω) (usually closed, that is, compact without boundary; an example nonclosed case is the cotangent bundle of a closed Riemannian manifold with the zero section as the Lagrangian submanifold) and a closed Lagrangian submanifold $L \subset M$, the problem (1–1) has been widely studied. The multiplicity problems of Hamiltonian systems on a symplectic manifold with Lagrangain boundary values are related to Arnold's conjecture about Lagrangian intersections. The autonomous case of this problem in $\mathbb{R}^{2n}$ is related to the Arnold chord conjecture. Generally, a Hamiltonian flow starting from a Lagrangian submanifold does not necessary return to the Lagrangian submanifold again. Arnold conjectured that, under some conditions, the Lagrangian intersection number has a lower bound estimated by the sum of all Beti numbers of the Lagrangian submanifold in the nondegenerate case; this sum is in turn estimated by the cup-length of the Lagrangian submanifold (see for example [Conley and Zehnder 1984; Hofer 1988; Floer 1988, 1989; Oh 1995; Ono 1996; Chekanov 1996, 1998; Liu 2005a]). For the Arnold chord conjecture, we mention [Arnold 1986; Mohnke 2001]. The multiplicity of the fixed energy problem (1–1) was studied in [Guo and Liu 2007]. The main differences between this work and the others are that here the symplectic manifold and the Lagrangian submanifold are not compact and all the topological data of the Lagrangian submanifold are trivial.

2. Some further properties of the Maslov type index theory

Liu [2007] developed some important properties of the L-index theory. In this section we study the relation between the L-index of solutions of Hamiltonian systems with L-boundary conditions and the Morse index of the corresponding functional defined via the Galerkin approximation method on the finite-dimensional truncated space at its corresponding critical points. Fei and Qiu [1996] treated the periodic case.

The eigenspace E_k of the operator $A = -J d/dt$ in the domain

$$W^{1,2}_{L_0}([0, 1],\ \mathbb{R}^{2n}) := \{z \in W^{1,2}([0, 1], \mathbb{R}^{2n}) : z(0) \in L_0,\ z(1) \in L_0\}$$

can be written as

$$E_k = -J \exp(k\pi t J)a_k = -J(\cos(k\pi t)I_{2n} + J \sin(k\pi t))a_k,$$
$$a_k = (a_{k1}, \cdots, a_{kn}, 0, \cdots, 0) \in \mathbb{R}^{2n}.$$

We define a Hilbert space

$$\mathrm{W}_{L_0} = W^{1/2,2}_{L_0}([0, 1], \mathbb{R}^{2n}) \subset \bigoplus_{k\in\mathbb{Z}} E_k$$

with L_0 boundary conditions

$$\mathrm{W}_{L_0} = \left\{ z \in L^2 \;\middle|\; z(t) = \sum_{k\in\mathbb{Z}} -J \exp(k\pi t J)a_k,\ \|z\|^2 := \sum_{k\in\mathbb{Z}} (1 + |k|)|a_k|^2 < \infty \right\}.$$

We denote its inner product by $\langle \cdot, \cdot \rangle$. By the well-known Sobolev embedding theorem, for any $s \in [1, +\infty)$, there is a constant $C_s > 0$ such that

$$\|z\|_{L^s} \le C_s \|z\| \quad \text{for all } z \in \mathrm{W}_{L_0}.$$

For any Lagrangian subspace $L \in \Lambda(n)$, suppose $P \in \mathrm{Sp}(2n) \cap O(2n)$ such that $L = PL_0$. Then we define $\mathrm{W}_L = P\mathrm{W}_{L_0}$. We denote by

$$\mathrm{W}^m_{L_0} = \bigoplus_{k=-m}^{m} E_k = \left\{ z \;\middle|\; z(t) = \sum_{k=-m}^{m} -J \exp(k\pi t J)a_k \right\}$$

the finite dimensional truncation of W_{L_0}, and $\mathrm{W}^m_L = P\mathrm{W}^m_{L_0}$.

Let $P^m = P^m_L : \mathrm{W}_L \to \mathrm{W}^m_L$ be the orthogonal projection for $m \in \mathbb{N}$. Then $\Gamma = \{P^m;\ m \in \mathbb{N}\}$ is a Galerkin approximation scheme with respect to A defined in (2–2) below, that is,

$$P^m \to I \text{ strongly as } m \to \infty \quad \text{and} P^m A = AP^m.$$

In this section we still consider the problem (1–1), with H satisfying

$$|H''(t, z)| \le a(1 + |z|^p) \quad \text{for all } (t, z) \in \mathbb{R} \times \mathbb{R}^{2n} \tag{2–1}$$

and for some $a > 0$, $p > 1$. We consider the functional on W_L

$$f(z) = \int_0^1 \big(\tfrac{1}{2}(-J\dot{z}, z) - H(t, z)\big)\, dt = \tfrac{1}{2}\langle Az, z\rangle - g(z), \quad z \in \mathrm{W}_L. \tag{2–2}$$

A critical point of f on W_L is a solution of (1–1). For a critical point $z = z(t)$, we denote $B(t) = H''(t, z(t))$ and define an operator B on W_L by

$$\langle Bz, w\rangle = \int_0^1 (B(t)z, w)dt.$$

Using the Floquet theory we have

$$\nu_L(B) = \dim\ker(A - B). \tag{2–3}$$

For $\delta > 0$, we denote by $m_\delta^*(\,\cdot\,)$, where $* = +, 0, -$, the dimension of the total eigenspace corresponding to the eigenvalue λ belonging to $[\delta, +\infty)$, $(-\delta, \delta)$, $(-\infty, -\delta]$, respectively, and denote by $m^*(\,\cdot\,)$, where again $* = +, 0, -$ the dimension of the total eigenspace corresponding to the eigenvalue λ belonging to $(0, +\infty)$, $\{0\}$, $(-\infty, 0)$, respectively. For any adjoint operator L, we define $L^\sharp = (L|_{ImL})^{-1}$, and we also define $P^m L P^m = (P^m L P^m)|_{\mathrm{W}_L^m}$. The following result is adapted from [Fei and Qiu 1996], where the periodic boundary condition was considered (see also [Long 1993]).

Theorem 2.1. *For any $B(t) \in C([0, 1], \mathrm{L}_s(\mathbb{R}^{2n}))$ having the pair of L indexes $(i_L(B), \nu_L(B))$ and any constant $0 < \delta \leq \frac{1}{4}\|(A - B)^\sharp\|$, there exists $m_0 > 0$ such that for $m \geq m_0$, we have*

$$\begin{aligned} m_\delta^+(P^m(A - B)P^m) &= mn - i_L(B) - \nu_L(B),\\ m_d^-(P^m(A - B)P^m) &= mn + i_L(B) + n,\\ m_\delta^0(P^m(A - B)P^m) &= \nu_L(B). \end{aligned} \tag{2–4}$$

Proof. We follow the ideas of [Fei and Qiu 1996].

Step 1. There is an $m_1 > 0$ such that for $m \geq m_1$

$$\dim\ker(P^m(A - B)P^m) \leq \dim\ker(A - B). \tag{2–5}$$

In fact, by contradiction it is easy to show that there is a constant $m_2 > 0$ such that for $m \geq m_2$

$$\dim P^m \ker(A - B) = \dim\ker(A - B). \tag{2–6}$$

Since B is compact, there is $m_1 \geq m_2$ such that for $m \geq m_1$

$$\|(I - P^m)B\| \leq 2\delta.$$

Take $m \geq m_1$, and let $\mathrm{W}_L^m = P^m \ker(A-B) \oplus Y^m$. Then $Y^m \subset \mathrm{Im}(A-B)$. For $y \in Y^m$ we have

$$y = (A-B)^\sharp (A-B) y = (A-B)^\sharp (P^m (A-B) P^m y + (P^m - I) B y).$$

This implies

$$\|y\| \leq \frac{1}{2\delta} \left\| P^m (A-B) P^m y \right\| \quad \text{for all } y \in Y^m. \tag{2–7}$$

By (2–6) and (2–7) we have (2–5).

Step 2. We distinguish two cases.

<u>Case 1: $\nu_L(B) = 0$.</u> By (2–3) and step 1 we obtain for $m \geq m_1$ that

$$m^0(P^m(A-B)P^m) = \dim\ker(A-B) = 0.$$

Since B is compact, there exists $m_3 \geq m_1$ such that, for $m \geq m_3$,

$$\|(I - P^m)B\| \leq \tfrac{1}{2} \|(A-B)^\sharp\|^{-1}.$$

Then $P^m(A-B)P^m = (A-B)P^m + (I-P^m)BP^m$ implies that

$$\|P^m(A-B)P^m z\| \geq \tfrac{1}{2} \|(A-B)^\sharp\|^{-1} \|z\| \quad \text{for all } z \in \mathrm{W}_L^m.$$

Thus the eigen-subspace $M_\delta^*(P^m(A-B)P^m)$ with eigenvalue λ belonging to the intervals $m_\delta^*(P^m(A-B)P^m)$ and the eigen-subspace $M^*(P^m(A-B)P^m)$ satisfy

$$M_\delta^*(P^m(A-B)P^m) = M^*(P^m(A-B)P^m) \quad \text{for } * = +, 0, -.$$

By Equation (A.5), there is $m_0 \geq m_3$ such that for $m \geq m_0$ the relation (2–4) holds.

<u>Case 2: $\nu_L(B) > 0$.</u> By step 1, it is easy to show that there exists $m_4 > 0$ such that for $m \geq m_4$

$$m_\delta^0(P^m(A-B)P^m) \leq \nu_L(B). \tag{2–8}$$

Let $\gamma \in \mathrm{P}(2n)$ be the fundamental solution of the linear Hamiltonian system

$$\dot{z} = JB(t)z.$$

Let γ_s, $0 \leq s \leq 1$ be the perturbed path defined by Equation (A.4). Define

$$B_s(t) = -J\dot{\gamma}_s(t)\gamma_s(t)^{-1}, \ t \in [0, 1].$$

Let B_s be the compact operator defined as B corresponding to $B_s(t)$. For $s \neq 0$, there holds $m^0(A - B_s) = 0$ and $\|B_s - B\| \to 0$ as $s \to 0$. If $s \in (0, 1]$, we have

$$i_L(\gamma_s) - i_L(\gamma_{-s}) = \nu_L(\gamma) = \nu_L(B), \ i_L(\gamma_{-s}) = i_L(B) = i_L(\gamma). \tag{2–9}$$

Choose $0 < s < 1$ such that $\|B - B_{\pm s}\| \leq \delta/2$. By case 1, (2–8), (2–9) and that

$$P^m(A - B_\pm)P^m = P^m(A - B)P^m + P^m(B - B_\pm)P^m,$$

there exists $m_0 \geq m_4$ such that for $m \geq m_0$

$$\begin{aligned} m_\delta^+(P^m(A-B)P^m) &\leq m^+(P^m(A-B_s)P^m) = mn - i_L(B) - \nu_L(B), \\ m_\delta^+(P^m(A-B)P^m) &\geq m^+(P^m(A-B_{-s})P^m) - m_\delta^0(P^m(A-B)P^m) \\ &\geq mn - i_L(B) - \nu_L(B). \end{aligned}$$

Hence, $m_\delta^0(P^m(A-B)P^m) = \nu_L(B)$ and

$$m_\delta^+(P^m(A-B)P^m) = mn - i_L(B) - \nu_L(B).$$

Note that $\dim \mathrm{W}_L^m = (2m+1)n$, so

$$m_\delta^-(P^m(A-B)P^m) = mn + n + i_L(B). \qquad \square$$

Corollary 2.2. *Let* $B_j(t) \in C([0,1], \mathrm{L}_s(\mathbb{R}^{2n}))$, $j = 1, 2$. *Assume* $B_1(t) < B_2(t)$, *that is*, $B_2(t) - B_1(t)$ *is positive definite for all* $t \in [0, 1]$. *Then there holds*

$$i_L(B_1) + \nu_L(B_1) \leq i_L(B_2).$$

Proof. Just as in Theorem 2.1, corresponding to $B_j(t)$ we have the operator B_j. Let $\Gamma = \{P^m\}$ be the approximation scheme with respect to the operator A. Then by (2–4), there exists $m_0 > 0$ such that if $m \geq m_0$ there holds

$$\begin{aligned} m_\delta^-(P^m(A-B_1)P^m) &= mn + n + i_L(B_1), \\ m_\delta^-(P^m(A-B_2)P^m) &= mn + n + i_L(B_2), \end{aligned}$$

where we choose $0 < \delta < \|B_2 - B_1\|/2$. Since $A - B_2 = (A - B_1) - (B_2 - B_1)$ and $B_2 - B_1$ is positive definite in $\mathrm{W}_L^m = P^m \mathrm{W}_L$ and $\langle (B_2 - B_1)x, x\rangle \geq 2\delta\|x\|$, we have $\langle (P^m(A - B_2)P^m)x, x\rangle \leq -\delta\|x\|$ with

$$x \in M_\delta^-(P^m(A-B_1)P^m) \oplus M_\delta^0(P^m(A-B_1)P^m).$$

This implies that $mn + n + i_L(B_1) + \nu_L(B_1) \leq mn + n + i_L(B_2)$. $\square$

Remark. From the proof of Corollary 2.2, it is easy to show that if $B_1(t) \leq B_2(t)$ for all $0 \leq t \leq 1$,

$$i_L(B_1) \leq i_L(B_2), \quad i_L(B_1) + \nu_L(B_1) \leq i_L(B_2) + \nu_L(B_2).$$

Definition 2.3. For any two matrix functions $B_j \in C([0,1], \mathrm{L}_s(\mathbb{R}^{2n}))$, $j = 0, 1$ with $B_0(t) < B_1(t)$ for all $t \in \mathbb{R}$, we define

$$I_L(B_0, B_1) = \sum_{s \in [0,1)} \nu_L((1-s)B_0 + sB_1).$$

Theorem 2.4. *For any two matrix functions $B_j \in C([0,1], \mathrm{L}_s(\mathbb{R}^{2n}))$ with $B_0(t) < B_1(t)$ for all $t \in \mathbb{R}$, we have*

$$I_L(B_0, B_1) = i_L(B_1) - i_L(B_0). \tag{2–10}$$

So we call $I_L(B_0, B_1)$ the relative L-index of the pair (B_0, B_1).

Proof. Step 1. By Corollary 2.2, if we denote $i_L(\lambda) = i_L((1-\lambda)B_0 + \lambda B_1)$, $\nu_L(\lambda) = \nu_L((1-\lambda)B_0 + \lambda B_1)$, there holds

$$i_L(\lambda_2) \geq i_L(\lambda_1) + \nu_L(\lambda_1), \text{ for } \lambda_2 > \lambda_1. \tag{2–11}$$

So the function $i_L(\lambda)$ is a monotone function in $[0,1]$.

Step 2. We prove that for any $\lambda \in [0,1)$ there holds

$$i_L(\lambda+0) = i_L(\lambda) + \nu_L(\lambda),$$

where $i_L(\lambda+0)$ is the right limit of $i_L(s)$ at λ. In fact, by (2–11), we have $i_L(\lambda) + \nu_L(\lambda) \leq i_L(\lambda+0)$. We now use the saddle point reduction methods to prove the opposite inequality $i_L(\lambda)+\nu_L(\lambda) \geq i_L(\lambda+0)$. Define $B_\lambda(t) = (1-\lambda)B_0(t) + \lambda B_1(t)$. We define in $L^2([0,1], \mathbb{R}^{2n})$

$$f_\lambda(x) = \int_0^1 [(-J\dot{x}(t), x(t)) - (B_\lambda(t)x(t), x(t))]\, dt \quad \text{for all } x \in \mathrm{dom}(A) = W_L.$$

Then by the saddle point reduction methods (see Equation (A.5)), we can reduce the functional f_λ in $L^2([0,1], \mathbb{R}^{2n})$ to a finite-dimensional subspace X of $L^2([0,1], \mathbb{R}^{2n})$ by $a_\lambda(x) = f_\lambda(u_\lambda(x))$, where $u_\lambda : X \to L^2([0,1], \mathbb{R}^{2n})$ is injective, and a_λ is continuous in λ. Denote the Morse indices of a_λ on X at $x = 0$ by m_λ^-, m_λ^0 and m_λ^-. If $\dim X = 2d + n$ large enough, we have from (A.5)

$$m_\lambda^- = d + n + i_L(\lambda), \quad m_\lambda^0 = \nu_L(\lambda), \quad m_\lambda^+ = d - i_L(\lambda) - \nu_L(\lambda). \tag{2–12}$$

For any fixed $\lambda \in [0,1)$, choosing $\mu \in (\lambda, 1) \cup [0, \lambda)$ sufficiently close to λ, we obtain

$$m_\lambda^\pm \leq m_\mu^\pm \leq m_\lambda^\pm + \nu_L(\lambda).$$

Then by (2–12), we have $i_L(\lambda) \leq i_L(\mu)$ and $i_L(\lambda) + \nu_L(\lambda) \geq i_L(\mu)$. This implies $i_L(\lambda) + \nu_L(\lambda) \geq i_L(\lambda+0)$ and $i_L(\lambda) \leq i_L(\lambda-0)$. But by (2–11), we have $i_L(\lambda) \geq i_L(\lambda-0)$, so $i_L(\lambda) = i_L(\lambda-0)$. That is to say, the function $i_L(\lambda)$ is left continuous at $(0,1]$. Moreover if $m_\lambda^0 = m^0$ is constant in some interval $[\lambda_1, \lambda_2]$, then $m_\lambda^- = m^-$ and $m_\lambda^+ = m^+$ are constant in this interval. Thus the function $i_L(\lambda)$ is locally constant at its continuous points, its discontinuous points are those with with $\nu_L(\lambda) > 0$, and there holds

$$i_L(1) = i_L(0) + \sum_{0 \leq \lambda < 1} \nu_L(\lambda),$$

which is exactly (2–10). □

Corollary 2.5. *If $\gamma \in \mathrm{P}(2n)$ is the fundamental solution of the linear Hamiltonian system with respect to $B(t) > 0$, there holds*

$$i_L(\gamma) = \sum_{0<t<1} \dim(\gamma(t)L \cap L). \tag{2–13}$$

Thus we can understand the index $i_L(\gamma)$ as a kind of intersection number of the two Lagrangian paths $w(t) = \gamma(t)L$ and $w_0(t) = L$.

Proof. We take $B_1(t) = B(t)$ and $B_0(t) = 0$ in Theorem 2.4. We note that the fundamental solution corresponding to $B_0(t) = 0$ is the constant path I. We have

$$I_L(0, B) = i_L(\gamma) - i_L(I).$$

But $i_L(I) = i_{L_0}(I) = -n$ and $B_s(t) = (1-s)B_0(t) + sB_1(t) = sB(t)$. The corresponding fundamental solution corresponding to $B_s(t) = sB(t)$ is $\gamma(st)$. Thus

$$I_L(0, B) = \sum_{s\in[0,1)} \nu_L(sB) = \sum_{s\in[0,1)} \dim[(\gamma(s)L) \cap L].$$

But $\dim[(\gamma(0)L) \cap L] = \dim L = n$, so we have (2–13). □

3. Dual index theory for linear Hamiltonian systems

Let $B \in C([0, 1], \mathrm{L}_s(\mathbb{R}^{2n}))$. Recall that $\mathrm{L}_s(\mathbb{R}^{2n})$ is the set of symmetric $2n \times 2n$ metrics. Consider the linear Hamiltonian system

$$\dot{z} = JB(t)z, \quad z \in \mathbb{R}^{2n}. \tag{3–1}$$

We consider in this section the dual Morse index theory of system (3–1) with Lagrangian boundary condition. The dual Morse index theory for periodic boundary condition was studied by Girardi and Matzeu [1991] for the cases of superquadatic Hamiltonian systems, and by the author in [Liu 2001] for the subquadratic Hamiltonian systems. This theory is an application of the Morse–Ekeland index theory [Ekeland 1990]. The dual action principal in Hamiltonian framework was first established by Clarke [1978; 1979; 1981] and Clarke and Ekeland [1978; 1980], and has since been adapted by many mathematicians to the study of various variational problems. The index theory for convex Hamiltonian systems was established by I. Ekeland (see for example [1990]), whose works are of fundamental importance in the study of convex Hamiltonian systems.

Let W_L be the Hilbert space defined by

$$W_L = \{z = (x, y)^T \in W^{1,2}([0, 1], \mathbb{R}^{2n}) |\, z(0),\ z(1) \in L\} \subset L^2.$$

The embedding $j : W_L \to \mathrm{L} = L^2([0,1], \mathbb{R}^{2n})$ is compact. Denote by $\langle \cdot, \cdot \rangle$ and $\langle \cdot, \cdot \rangle_2$ the respective inner products on W_L and L. We define an operator $A : \mathrm{L} \to \mathrm{L}$ with domain W_L by $A = -J d/dt$. The spectrum of A is isolated, and in fact, $\sigma(A) = \pi\mathbb{Z}$. Let $k \notin \sigma(A)$ be so large such that $B(t) + kI > 0$. Then the operator $\Lambda_k = A + kI : W_L \to \mathrm{L}$ is invertible, and its inverse is compact. We define a quadratic form in L by

$$Q^*_{k,B}(v, u) = \int_0^1 \big((C_k(t)v(t), u(t)) - (\Lambda_k^{-1} v(t), u(t))\big)\, dt \quad \text{for all } v, u \in \mathrm{L},$$

where $C_k(t) = (B(t) + kI)^{-1}$. Define $Q^*_{k,B}(v) = Q^*_{k,B}(v, v)$. Then

$$\langle C_k v, v \rangle_2 = \int_0^1 (C_k(t)v(t), v(t))\, dt$$

defines a Hilbert structure in L. $C_k^{-1}\Lambda_k^{-1}$ is a self-adjoint and compact operator under this inner product. By the spectral theory, there exists a basis e_j, $j \in \mathbb{N}$ of L, and an eigenvalue sequence $\lambda_j \to 0$ in $\mathbb{R}$ such that

$$\begin{aligned} &\langle C_k e_i, e_j \rangle_2 = \delta_{ij}, \\ &\langle \Lambda_k^{-1} e_j, v \rangle_2 = \langle C_k \lambda_j e_j, v \rangle_2 \quad \text{for all } v \in \mathrm{L}. \end{aligned}$$

For any $v \in \mathrm{L}$ with $v = \sum_{j=1}^{\infty} \xi_j e_j$, there holds

$$Q^*_{k,B}(v) = -\int_0^1 (\Lambda_k^{-1} v(t), v(t)) - (C_k(t)v(t), v(t))\, dt = \sum_{j=1}^{\infty} (1 - \lambda_j)\xi_j^2.$$

Define

$$\begin{aligned} \mathrm{L}_k^-(B) &= \Big\{ \sum_{j=1}^{\infty} \xi_j e_j \;\Big|\; \xi_j = 0 \text{ if } 1 - \lambda_j \ge 0 \Big\}, \\ \mathrm{L}_k^0(B) &= \Big\{ \sum_{j=1}^{\infty} \xi_j e_j \;\Big|\; \xi_j = 0 \text{ if } 1 - \lambda_j \ne 0 \Big\}, \\ \mathrm{L}_k^+(B) &= \Big\{ \sum_{j=1}^{\infty} \xi_j e_j \;\Big|\; \xi_j = 0 \text{ if } 1 - \lambda_j \le 0 \Big\}. \end{aligned}$$

Observe that $\mathrm{L}_k^-(B)$, $\mathrm{L}_k^0(B)$ and $\mathrm{L}_k^+(B)$ are $Q^*_{k,B}$-orthogonal, and also that $\mathrm{L} = \mathrm{L}_k^-(B) \oplus \mathrm{L}_k^0(B) \oplus \mathrm{L}_k^+(B)$. Since $\lambda_j \to 0$ as $j \to \infty$, both $\mathrm{L}_k^-(B)$ and $\mathrm{L}_k^0(B)$ are finite subspaces. We define the k-dual Morse index of B by

$$i_k^*(B) = \dim \mathrm{L}_k^-(B), \qquad \nu_k^*(B) = \dim \mathrm{L}_k^0(B).$$

Theorem 3.1. *There holds*

$$i_k^*(B) = i_L(B) + n + n\left[\frac{k}{\pi}\right], \quad \nu_k^*(B) = \nu_L(B), \tag{3–2}$$

where $[a] = \max\{j \in \mathbb{Z} \mid j \le a\}$.

Proof. We only prove (3–2) for the special case $L = L_0$. We first define a functional on

$$W^m = \left\{ x \;\middle|\; x(t) = \sum_{j=-m}^{m} -J \exp(j\pi t J) a_j,\ a_j \in \mathbb{R}^n \oplus \{0\} \subset \mathbb{R}^{2n} \right\}$$

by

$$\begin{aligned} Q_m(x) &= \int_0^1 [(\Lambda_k x(t), x(t)) - (C_k^{-1}(t)x, x)]\, dt \\ &= \int_0^1 [(-J\dot{x}(t), x(t)) - (B(t)x(t), x(t))]\, dt \quad \text{for all } x \in W^m. \end{aligned}$$

We define two linear operators A_k and B_k from W^m onto its dual space $W^{m*} \cong W^m$ such that

$$\begin{aligned} \langle A_k x, y\rangle_2 &= \int_0^1 (\Lambda_k x(t), y(t))\, dt && \text{for all } x, y \in W^m, \\ \langle B_k x, y\rangle_2 &= \int_0^1 ((B(t) + kI)x(t), y(t))\, dt && \text{for all } x, y \in W^m. \end{aligned}$$

Next $\langle \cdot, \cdot \rangle_m := \langle B_k \cdot, \cdot \rangle_2$ is a inner product in W^m. We consider the eigenvalues $\mu_j \in \mathbb{R}$ of A_k with respect to this inner product, that is,

$$A_k x_j = \mu_j B_k x_j$$

for some $x_j \in W^m \setminus \{0\}$. Suppose $\mu_1 \le \mu_2 \le \cdots \le \mu_l$ with $l = \dim W^m = 2mn + n$ (each eigenvalue is counted with its multiplicity), and construct a basis in W^m of eigenvectors $v_1, \dots, v_l$ such that, for $i, j = 1, 2, \dots, l$,

$$\begin{aligned} \langle v_i, v_j \rangle_m &= \delta_{ij}, \\ \langle A_m v_i, v_j \rangle_m &= \mu_i \delta_{ij}, \\ Q_m(v_i, v_j) &= (\mu_i - 1)\delta_{ij}. \end{aligned}$$

The Morse indexes $m^-(Q_m)$, $m^0(Q_m)$ and $m^+(Q_m)$ of Q_m satisfy

$$\begin{aligned} m^-(Q_m) &=^\sharp \{\mu_j \mid 1 \le j \le l,\ \mu_j < 1\}, \\ m^+(Q_m) &=^\sharp \{\mu_j \mid 1 \le j \le l,\ \mu_j > 1\}, \\ m^0(Q_m) &=^\sharp \{\mu_j \mid 1 \le j \le l,\ \mu_j = 1\}. \end{aligned}$$

By Theorem 2.1, we have for $m > 0$ large enough

$$m^-(Q_m) = mn + n + i_L(B), \qquad m^0(Q_m) = \nu_L(B). \tag{3–3}$$

We denote by $Q^*_{k,m}$ the restriction of the quadratic Q^*_k to the subspace W^m, and define $i^*_{k,m}(B) = m^-(Q^*_{k,m})$, $\nu^*_{k,m}(B) = m^0(Q^*_{k,m})$. By an argument from [Girardi and Matzeu 1991], we have $i^*_{k,m}(B) \to i^*_k(B)$ and $\nu^*_{k,m}(B) \to \nu^*_k(B)$ as $m \to \infty$. Let $v'_j = A_m v_j$ for $j = 1, 2, \ldots, l$. It is a basis of W^m and

$$Q^*_{k,m}(v'_i, v'_j) = \begin{cases} 0, & \text{for } i \neq j, \\ \mu_j(\mu_j - 1), & \text{for } i = j. \end{cases}$$

$Q^*_{k,m}(v'_j)$ is negative if and only if $0 < \mu_j < 1$. We now deduce the total multiplicity of the negative eigenvalues $\mu_j < 0$. If one replaces the inner product $\langle \cdot, \cdot \rangle_m$ by the usual one, that is , one replaces the matrix B_k by the identity I, the eigenvalues μ_j should be replaced by the eigenvalues η_j of A_m with respect to the standard inner product. It is easy to check that μ_j and η_j have the same signs. So the total multiplicity of negative μ_j's equals the total multiplicity of negative η_h's. But we have

$$\eta_h = h\pi + k, \quad -m \leq h \leq m,$$

and each has multiplicity n. Therefore, the total multiplicity of the negative η_h is $n(m - [k/\pi])$. So the total multiplicity of $\mu_j \in (0, 1)$ is $m^-(Q_m) - n(m - [k/\pi])$. By definition we have

$$i^*_{k,m}(B) = m^-(Q_m) - n(m - [k/\pi]).$$

So for $m > 0$ large enough, from (3–3) we get (3–2). □

Corollary 3.2. 3.2 *Under the condition of Equation* (2–3), *there holds*

$$I_L(B_0, B_1) = i^*_k(B_1) - i^*_k(B_0).$$

4. Proof of Theorem 1.1 and some consequences

Lemma 4.1 [Chang 1981, Theorem 5.1, Corollary II.5.2]. *Let $f \in C^2(\mathrm{L}, \mathbb{R})$ satisfy the* (*PS*) *condition $f'(0) = 0$ and suppose there exists*

$$r \notin [m^-(f''(0)), m^-(f''(0)) + m^0(f''(0))]$$

with $H_q(\mathrm{L}, f_a; \mathbb{R}) \cong \delta_{q,r}\mathbb{R}$. Then f has at least one nontrivial critical point $u_1 \neq 0$. Moreover, if $m^0(f''(0)) = 0$ and $m^0(f''(u_1)) \leq |r - m^-(f''(0))|$, then f has one more nontrivial critical point $u_2 \neq u_1$.

Theorem 1.1. Without loss any generality we can suppose $H(t, 0) = 0$ and $L = L_0$. By the condition (H_∞) and the remark after Equation (2–1), we get that $i_L(B_1) + \nu_L(B_1) \leq i_L(B_2) + \nu_L(B_2)$, and so we have $\nu_L(B_1) = 0$. We shall first prove that under the above conditions (1–2) or (1–3), there holds

$$i_L(B_1) \notin [i_L(B_0), i_L(B_0) + \nu_L(B_0)].$$

More clearly, under the condition (1–2), it is claimed

$$i_L(B_1) = i_L(B_1) + \nu_L(B_1) < i_L(B_0), \tag{4–1}$$

and under the condition (1–3), it is claimed

$$i_L(B_0) + \nu_L(B_0) < i_L(B_1). \tag{4–2}$$

We first prove (4–1). By Equation (2–1) and condition (1–2), we have

$$i_L(B_1) \le i_L(B_1 + kI) \le i_L(B_0).$$

We shall prove

$$i_L(B_1) < i_L(B_1 + kI).$$

In fact, suppose

$$\gamma_1(t) = \begin{pmatrix} S_1(t) & V_1(t) \\ T_1(t) & U_1(t) \end{pmatrix} \in \mathrm{P}(2n)$$

is a symplectic path that is the fundamental solution of the linear Hamiltonian system associated with the matrix function $B_1(t)$. Since $JB_1(t) = B_1(t)J$, one can show that $\exp(Jkt)\gamma_1(t)$ is the fundamental solution of the linear Hamiltonian system

$$\dot{z} = J(B_1(t) + kI)z.$$

One has

$$\exp(Jkt)\gamma_1(t) = \begin{pmatrix} S_1(t)\cos kt - T_1(t)\sin kt & V_1(t)\cos kt - U_1(t)\sin kt \\ S_1(t)\sin kt + T_1(t)\cos kt & V_1(t)\sin kt + U_1(t)\cos kt \end{pmatrix}.$$

The associated unitary $n \times n$ matrix $\mathrm{Q}(t)$ defined by (2–2) with respect to the above matrix is

$$\begin{aligned} \mathrm{Q}(t) &= [U_1(t) - \sqrt{-1}V_1(t)][U_1(t) + \sqrt{-1}V_1(t)]^{-1}\exp(2k\sqrt{-1}t) \\ &= \mathrm{Q}_1(t)\exp(2k\sqrt{-1}t). \end{aligned}$$

In Equation (A.6), $\Delta_j = \theta_j(1) - \theta_j(0)$ and $\Delta_j^1 = \theta_j^1(1) - \theta_j^1(0)$, associated respectively to $\mathrm{Q}(t)$ and $\mathrm{Q}_1(t)$, satisfy

$$\Delta_j = \theta_j(1) - \theta_j(0) = \Delta_j^1 + 2k = \theta_j^1(1) - \theta_j^1(0) + 2k.$$

Since $k \ge \pi$, there holds

$$i_L(B_1) + n \le i_L(B_1 + kI). \tag{4–3}$$

Thus we have proved (4–1), and (4–2) can be proved similarly.

By the condition (H_∞), $H''(t,x)$ is bounded and there exist $\mu_1,\ \mu > 0$ such that

$$I \le H''(t,x) + \mu I \le \mu_1 I \quad \text{for all } (t,x). \tag{4–4}$$

We define a convex function $N(t,x)=H(t,x)+\mu|x|^2/2$. Its Fenchel dual defined by

$$N^*(t,x)=\sup_{y\in\mathbb{R}^{2n}}\{(x,y)-N(t,y)\}$$

satisfies (see [Ekeland 1990])

$$N^*\in C^2([0,1]\times\mathbb{R}^{2n},\mathbb{R}),$$
$$N^{*\prime\prime}(t,y)=N''(t,x)^{-1}\quad\text{for } y=N'(t,x).$$

From (4–4) we have

$$\mu_1^{-1}I\le N^{*\prime\prime}(t,y)\le I\quad\text{for all }(t,y). \tag{4–5}$$

So we have $|x|\to\infty$ if and only if $|y|\to\infty$ with $y=N'(t,x)$. Thus there exists $r_1>0$ such that

$$(B_2(t)+\mu I)^{-1}\le N^{*\prime\prime}(t,y)\le(B_1(t)+\mu I)^{-1} \tag{4–6}$$

for all t,y with $|y|\ge r_1$. We choose $\mu>0$ satisfying (4–4) and $\mu\notin\sigma(A)$. We recall that $(\Lambda_\mu x)(t)=-J\dot{x}(t)+\mu x(t)$. We consider the functional

$$f(u)=-\frac{1}{2}\int_0^1\left[(\Lambda_\mu^{-1}u(t),u(t))-N^*(t,u(t))\right]dt\quad\text{for } u\in\mathrm{L}.$$

It is easy to see that $f\in C^2$ and satisfies (PS) condition (see [Ekeland 1990]). There is a one to one correspondence from the critical points of f to the solutions of Hamiltonian systems (1–1). We note that 0 is a trivial critical point of f and $N^{*\prime}(t,0)=0$. At every critical point u_0, the second variation of f defines a quadratic form on L by

$$(f''(u_0)u,u)=-\int_0^1\left[(\Lambda_\mu^{-1}u(t),u(t))-(N^{*\prime\prime}(t,u_0(t))u(t),u(t))\right]dt\quad\text{for } u\in\mathrm{L}.$$

Its Morse index and nullity are both finite we denote by $(i_\mu^*(u_0),\nu_\mu^*(u_0))$ the index pair. The critical point u_0 corresponds to a solution $x_0=\Lambda_\mu^{-1}u_0$ of (1–1), and $N^{*\prime\prime}(t,u_0(t))=N''(t,x_0(t))^{-1}$. So by Theorem 3.1, we have

$$i_\mu^*(u_0)=i_L(x_0)+n+n\left[\frac{\mu}{\pi}\right],\ \nu_\mu^*(u_0)=\nu_L(x_0).$$

The index pair $(i_L(x_0),\nu_L(x_0))$ is the L-index of the linear Hamiltonian system

$$\dot{y}(t)=JH''(t,x_0(t))y(t).$$

By condition (1–2) and the result (4–3), we have

$$i_L(B_1)+\nu_L(B_1)+n\le i_L(B_0). \tag{4–7}$$

By condition (1–3), similarly we have

$$i_L(B_0)+\nu_L(B_0)+n \le i_L(B_1).$$

From (4–7) and the above inequality, we have that

$$|i_L(B_0)-i_L(B_1)| \ge n \quad \text{and} \quad |i^*_\mu(B_0)-i^*_\mu(B_1)| \ge n. \tag{4–8}$$

In the following, we need to prove that the homology groups satisfy

$$H_q(\mathrm{L},\, f_a;\mathbb{R}) \cong \delta_{qr}\mathbb{R}, \quad q=0,1,\dots, \tag{4–9}$$

for some $a\in\mathbb{R}$ and $r=i^*_\mu(B_1)$. $f_a=\{x\in \mathrm{L}\mid f(x)\le a\}$ is the level set below a. We follow the ideas of the proof of Lemma II.5.1 in [Chang 1981] to prove (4–9). See [Dong 2005] and [Liu 2005b] for some similar computations.

Step 1. Under the condition (H_∞), there holds

$$\mathrm{L}=\mathrm{L}^-_\mu(B_1)\oplus \mathrm{L}^+_\mu(B_2),$$

where $\mathrm{L}^*_\mu(B)$ for $*=\pm, 0$ is defined in Section 3. In fact, it is clear that $\mathrm{L}^-_\mu(B_1)\cap \mathrm{L}^+_\mu(B_2)=\{0\}$. By $\nu^*_\mu(B_2)=\nu_L(B_2)=0$, we have $\mathrm{L}=\mathrm{L}^-_\mu(B_2)\oplus\mathrm{L}^+_\mu(B_2)$. By condition (H_∞), we have $i^*_\mu(B_1)=i^*_\mu(B_2)=r$. Suppose $\xi_1,\xi_2,\cdots,\xi_r$ is a basis in $\mathrm{L}^-_\mu(B_1)$. Decompose ξ_j by $\xi_j=\xi^-_j+\xi^+_j$ with $\xi_j\in\mathrm{L}^\pm_\mu(B_2)$. It is clear that $\xi^-_1,\cdots,\xi^-_r$ are linear independent, so it is a basis for $\mathrm{L}^-_\mu(B_2)$. For any $\xi\in\mathrm{L}$, there holds $\xi=\xi^-+\xi^+$ with $\xi^\pm\in\mathrm{L}^\pm_\mu(B_2)$. Suppose $\xi^-=a_1\xi^-_1+\cdots+a_r\xi^-_r$. Then

$$\xi=\sum_{j=1}^r a_j\xi_j+(\xi^+-\sum_{j=1}^r a_j\xi^+_j)=\xi_1+\xi_2$$

with $\xi_1\in\mathrm{L}^-_\mu(B_1)$ and $\xi_2\in\mathrm{L}^+_\mu(B_2)$.

Step 2. For sufficiently small $s>0$, from the structure of the symplectic group and the definition of the Maslov-type index, we know that $\nu_L(B_1-sI)=\nu_L(B_1)=0$, and $\nu_L(B_2+sI)=\nu_L(B_2)=0$, and so $i_L(B_1-sI)=i_L(B_1)=i_L(B_2)=i_L(B_2+sI)$. Denote the so-called deformation space by

$$D_R=\mathrm{L}^-_\mu(B_1-sI)\oplus\{u\in\mathrm{L}^+_\mu(B_2+sI)\mid \|u\|\le R\}.$$

For $R>0$ and $-a>0$ large, we have the deformation result

$$H_q(\mathrm{L},\, f_a;\mathbb{R})=H_q(D_R,\, D_R\cap f_a;\mathbb{R}). \tag{4–10}$$

The proof of (4–10) is standard in the Morse theory [Bott 1982]. We only need to use the negative flow to deform $(\mathrm{L},\, f_a)$ to $(D_R,\, D_R\cap f_a)$. For any $u=u_1+u_2\in\mathrm{L}$

with $u_1 \in \mathrm{L}_\mu^-(B_1 - sI)$ and $u_2 \in \mathrm{L}_\mu^+(B_2 + sI)$, by the self-adjointness, we have

$$\begin{aligned}(f'(u), u_2 - u_1) &= -\int_0^1 dt\, \big[(\Lambda^{-1}u, u_2 - u_1) - (N^{*\prime}(t, u), u_2 - u_1)\big]\\ &= \int_0^1 dt\, \big[(\Lambda^{-1}u_1, u_1) - (\Lambda^{-1}u_2, u_2)\big]\\ &\quad + \int_0^1 dt\, \Big(\int_0^1 d\tau\, N^{*\prime\prime}(t, \tau u)(u_1 + u_2), u_2 - u_1\Big)\\ &= \int_0^1 dt\, (\Lambda^{-1}u_1, u_1) - \int_0^1 dt\, \Big(\int_0^1 d\tau\, N^{*\prime\prime}(t, \tau u)u_1, u_1\Big)\\ &\quad - \int_0^1 dt\, (\Lambda^{-1}u_2, u_2) + \int_0^1 dt\, \Big(\int_0^1 d\tau\, N^{*\prime\prime}(t, \tau u)u_2, u_2\Big).\end{aligned}$$

By (4–5) and (4–6), we have

$$\begin{aligned}&\int_0^1 dt\, \Big(\int_0^1 d\tau\, N^{*\prime\prime}(t, \tau u)u_1, u_1\Big)\\ &\quad = \int_0^1 dt \int_0^{h(t,u)} d\tau\, (N^{*\prime\prime}(t, \tau u)u_1, u_1) + \int_0^1 dt \int_{h(t,u)}^1 d\tau\, (N^{*\prime\prime}(t, \tau u)u_1, u_1)\\ &\quad \le c_0\|u\| + \int_0^1 dt\, ((B_1(t) + \mu I - sI)u_1, u_1),\end{aligned}$$

where $h(t, u) = r_1/|u(t)|$. Similarly,

$$\begin{aligned}\int_0^1 dt\, \Big(\int_0^1 d\tau\, N^{*\prime\prime}(t, \tau u)u_2, u_2\Big) &\ge \int_0^1 dt \int_{h(t,u)}^1 d\tau\, (N^{*\prime\prime}(t, \tau u)u_2, u_2)\\ &\ge \int_0^1 dt\, ((B_2(t) + \mu I + sI)u_2, u_2) - c\|u\|\end{aligned}$$

for some $c > 0$. So by the last three relations, we have

$$(f'(u), u_2 - u_1) \ge c_1\|u_1\|^2 + c_2\|u_2\|^2 - c_3(\|u_1\| + \|u_2\|).$$

Thus for large R with $\|u_1\| \ge R$ or $\|u_2\| \ge R$, we have

$$(-f'(u), u_2 - u_1) < -1. \tag{4–11}$$

We know from (4–11) that f has no critical point outside D_R, and that $-f'(u)$ points inward to D_R on ∂D_R. So we can define the deformation by negative flow. In fact, for any $u = u_1 + u_2 \notin D_R$, let $\sigma(\theta, u) = e^\theta u_1 + e^{-\theta} u_2$, and $d_u = \log\|u_2\| - \log R$. We define the deformation map $\eta : [0, 1] \times \mathrm{L} \to \mathrm{L}$ by

$$\eta(\theta, u_1 + u_2) = \begin{cases} u_1 + u_2, & \|u_2\| \le R,\\ \sigma(d_u\theta, u), & \|u_2\| > R.\end{cases}$$

The map η satisfies the properties

$$\eta(0,\cdot)=\mathrm{id},\qquad \eta(1,\mathrm{L})\subset D_R,\qquad \eta(1,f_a)\subset D_R\cap f_a$$
$$\eta(\theta,f_a)\subset f_a,\quad \eta(\theta,\cdot)|_{D_R}=id|_{D_R}.$$

Thus the pair $(D_R, D_R\cap f_a)$ is a deformation retract of the pair (L, f_a).

Step 3. For large R, $-a>0$, there holds

$$H_q(D_R, D_R\cap f_a)\cong \delta_{q,r}\mathbb{R}.$$

In fact, similarly to the above computation, for large $m>0$, we have

$$\begin{aligned}
&\int_0^1 dt\, N^*(t,u(t))\\
&\quad=\int_0^1 dt\Big(N^*(t,0)+\iint\limits_{[0,1]\times[0,1]} d\tau\, ds\, \tau(N^{*\prime\prime}(t,\tau s u(t))u(t),u(t))\Big)\\
&\quad\le \int_{|u(t)|\ge m r_1} dt \iint\limits_{[0,1]\times[0,1]} d\tau\, ds\, \tau(N^{*\prime\prime}(t,\tau s u(t))u(t),u(t))+c_m\\
&\quad\le \int\limits_{|u(t)|\ge m r_1} dt \iint\limits_{|s\tau u(t)|\ge r_1,\ \tau,s\in[0,1]} d\tau\, ds\, \tau(N^{*\prime\prime}(t,\tau s u(t))u(t),u(t))\\
&\qquad\qquad+\int\limits_{|u(t)|\ge m r_1} dt \iint\limits_{|s\tau u(t)|\le r_1,\ \tau,s\in[0,1]} d\tau\, ds\, \tau(N^{*\prime\prime}(t,\tau s u(t))u(t),u(t))+c_m\\
&\quad\le \frac{1}{2}\int_0^1 dt\,((B_1(t)+\mu I)^{-1}u(t),u(t))+k_m\|u\|+c_m,
\end{aligned}$$

where c_m and k_m are constants depending only on m and $k_m\to 0$ as $m\to+\infty$. So for the small s in the step 2 above, we can choose a large number m such that

$$\int_0^1 dt\, N^*(t,u(t))\le \frac{1}{2}\int_0^1 dt\,((B_1(t)+\mu I-sI)^{-1}u(t),u(t))+C\quad\text{for all } u\in\mathrm{L}$$

for some constant $C>0$. Thus for any $u=u_1+u_2$ with $u_1\in \mathrm{L}^-_\mu(B_1-sI)$ and $u_2\in\mathrm{L}^+_\mu(B_2+sI)$ with $\|u_2\|\le R$, there holds

$$f(u)\le -C_1\|u_1\|^2+C_2\|u_1\|+C_3,$$

where $C_j,\ j=1,2,3$ are constants and $C_1>0$. It implies that $f(u)\to-\infty$ if and only if $\|u_1\|\to\infty$ uniformly for $u_2\in\mathrm{L}^+_\mu(B_2+sI)$ with $\|u_2\|\le R$. In the following we denote by $B_r=\{x\in\mathrm{L}|\ \|x\|\le r\}$ the ball with radius r in L. Therefore for $-a_1>-a_2$ sufficiently large, there exist three numbers with $R<R_1<R_2<R_3$

satisfying

$$\begin{aligned}&(\mathrm{L}_\mu^+(B_2+sI)\cap B_{R_3})\oplus(\mathrm{L}_\mu^-)(B_1-sI)\setminus B_{R_2})\subset f_{a_1}\cap D_{R_3}\\ &\subset(\mathrm{L}_\mu^+(B_2+sI)\cap B_{R_3})\oplus(\mathrm{L}_\mu^-)(B_1-sI)\setminus B_{R_1})\subset f_{a_2}\cap D_{R_3}.\end{aligned}$$

Recall that $\sigma(\theta,u)=e^\theta u_1+e^{-\theta}u_2$. By definition, we have $f(\sigma(0,u))=f(u)>a_1$ and $f(\sigma(\theta,u))\to-\infty$ as $\theta\to\infty$ if $u=u_1+u_2\in D_{R_3}\cap(f_{a_2}\setminus f_{a_1})$. It implies that there exists $\theta_0=\theta_0(u)>0$ such that $f(\sigma(\theta_0,u))=a_1$. But by (4–11),

$$\frac{d}{d\theta}f(\sigma(\theta,u))\le-1\quad\text{at any point }\theta>0.$$

By the implicit function theorem, $\theta_0(u)$ is continuous in u. We define another deformation map $\eta_0:[0,1]\times f_{a_2}\cap D_{R_3}\to f_{a_2}\cap D_{R_3}$ by

$$\eta_0(\theta,u)=\begin{cases}u & u\in f_{a_1}\cap D_{R_3},\\ \sigma(\theta_0(u)\theta,u), & u\in D_{R_3}\cap(f_{a_2}\setminus f_{a_1}).\end{cases}$$

It is clear that η_0 is a deformation from $f_{a_2}\cap D_{R_3}$ to $f_{a_1}\cap D_{R_3}$. We now define

$$\tilde\eta(u)=d(\eta_0(1,u))\quad\text{with }d(u)=\begin{cases}u, & \|u_1\|\ge R_1,\\ u_2+\dfrac{u_1}{\|u_1\|}R_1, & 0<\|u_1\|<R_1.\end{cases}$$

This map defines a strong deformation retract:

$$\tilde\eta:D_{R_3}\cap d_{a_2}\to\big(\mathrm{L}_\mu^+(B_2+sI)\cap B_{R_3}\big)\oplus\big(\mathrm{L}_\mu^-(B_1-sI)\cap\{u\in\mathrm{L}\mid\|u\|\ge R_1\}\big).$$

Now we can compute the homology groups

$$\begin{aligned}&H_q(D_{R_3},D_{R_3}\cap f_{a_2};\mathbb{R})\\ &\quad\cong H_q(D_{R_3},(\mathrm{L}_\mu^+(B_2+sI)\cap B_{R_3})\oplus(\mathrm{L}_\mu^-(B_1-sI)\cap\{u\in\mathrm{L}\mid\|u\|\ge R_1\});\mathbb{R})\\ &\quad\cong H_q(\mathrm{L}_\mu^-(B_1-sI)\cap B_{R_3},\partial(\mathrm{L}_\mu^-(B_1-sI)\cap B_{R_3});\mathbb{R})\\ &\quad\cong\delta_{qr}\mathbb{R}.\end{aligned}$$

From (4–8), (4–9), and (A.2) below, and by using Equation (4–1), we complete the proof. □

Corollary 4.2. *Let H satisfy the conditions* (H_0) *and* (H_∞), *and suppose* $B_0(t)=H''(t,0)$ *satisfies one of the twisted conditions*:

(i) $B_1(t)<B_0(t)$, *there exists* $\lambda\in(0,1)$ *such that* $\nu_L((1-\lambda)B_1+\lambda B_0)\ne0$;

(ii) $B_0(t)<B_1(t)$, *there exists* $\lambda\in(0,1)$ *such that* $\nu_L((1-\lambda)B_0+\lambda B_1)\ne0$.

Then (1–1) *possesses at least one nontrivial solution. Furthermore, if* $\nu_L(B_0)=0$ *and in* (*i*) *we replace the second condition by* $\sum_{\lambda\in(0,1)}\nu((1-\lambda)B_1+\lambda B_0)\ge n$, *or in* (*ii*) *we replace the second condition by* $\sum_{\lambda\in(0,1)}\nu((1-\lambda)B_0+\lambda B_1)\ge n$, *the Hamiltonian system* (1–1) *possesses at least two nontrivial solutions.*

Proof. It follows from (2–3), the proof of Theorem 1.1 and (4–2). In the first case, we have $r = i_L(B_1) \notin [i_L(B_0), i_L(B_0) + \nu_L(B_0)]$. In the second case we have $|i_L(B_0) - i_L(B_1)| \geq n$. □

The proof of Theorem 1.1 in fact proves this:

Theorem 4.3. *Let H satisfy conditions (H_0) and (H_∞). Suppose $B_0(t) = H''(t, 0)$ satisfies the twisted conditions*

$$i_L(B_1) \notin [i_L(B_0), i_L(B_0) + \nu_L(B_0)].$$

Then the problem (1–1) *possesses at least one nontrivial solution. Moreover, if $\nu_L(B_0) = 0$ and $|i_L(B_1) - i_L(B_0)| \geq n$, then* (1–1) *possesses at least two nontrivial solutions.*

Remark. The condition $B_1(t) < B_2(t)$ in Theorem 2.4 can be replaced by $B_1(t) \leq B_2(t)$ for all t and $B_2 - B_1 \geq \delta > 0$ for some constant δ as an operator in L. So the conditions in parts (i) and (ii) of Corollary 4.2 can be replaced by this kind of condition. The condition $JB_1(t) = B_1(t)J$ in (H_∞) can be replaced by $JB_0(t) = B_0(t)J$.

Appendix. Maslov-type index for symplectc paths with Lagrangian boundary condition

We give a brief introduction to the Maslov-type index for symplectc paths with Lagrangian boundary condition. The details can be found in [Liu 2007]. We denote the symplectic group by

$$\mathrm{Sp}(2n) = \left\{ M \in \mathrm{L}(\mathbb{R}^{2n}) \mid M^T J M = J \right\},$$

and denote the symplectic path space by

$$\mathrm{P}(2n) = \{ \gamma \in C([0, 1], \mathrm{Sp}(2n)) \mid \gamma(0) = I_{2n} \}.$$

We write a symplectic path $\gamma \in \mathrm{P}(2n)$, in the form

$$\gamma(t) = \begin{pmatrix} S(t) & V(t) \\ T(t) & U(t) \end{pmatrix}, \tag{A.1}$$

where $S(t), T(t), V(t), U(t)$ are $n \times n$ matrices. The n vectors coming from the rightmost columns of the above matrix are linearly independent and they span a Lagrangian subspace of $(\mathbb{R}^{2n}, \omega_0)$. In particular, at $t = 0$, this Lagrangian subspace is $L_0 = \{0\} \oplus \mathbb{R}^n$.

Definition A.1. We define the L_0-nullity of any symplectic path $\gamma \in \mathrm{P}(2n)$ by

$$\nu_{L_0}(\gamma) \equiv \dim \ker_{L_0}(\gamma(1)) := \dim \ker V(1) = n - \operatorname{rank} V(1) \tag{A.2}$$

with the $n \times n$ matrix function $V(t)$ defined in (A.1).

We define two subsets of $\mathrm{P}(2n)$ by

$$\mathrm{P}(2n)^*_{L_0} = \{\gamma \in \mathrm{P}(2n) \mid \nu_{L_0}(\gamma) = 0\},$$
$$\mathrm{P}(2n)^0_{L_0} = \{\gamma \in \mathrm{P}(2n) \mid \nu_{L_0}(\gamma) > 0\}.$$

We note that

$$\operatorname{rank}\begin{pmatrix} V(t) \\ U(t) \end{pmatrix} = n,$$

so the complex matrix $U(t) \pm \sqrt{-1}V(t)$ is invertible. We define a complex matrix function by

$$\mathrm{Q}(t) = \big(U(t) - \sqrt{-1}V(t)\big)\big(U(t) + \sqrt{-1}V(t)\big)^{-1}. \tag{A.3}$$

It is easy to see that the matrix $\mathrm{Q}(t)$ is a unitary matrix for any $t \in [0, 1]$. We define

$$M_+ = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}, \quad M_- = \begin{pmatrix} 0 & J_n \\ -J_n & 0 \end{pmatrix}, \quad J_n = \mathrm{diag}(-1, 1, \dots, 1).$$

For a path $\gamma \in \mathrm{P}(2n)^*_{L_0}$, we first adjoin it with a simple symplectic path starting from $J = -M_+$, that is, we define a symplectic path by

$$\tilde{\gamma}(t) = \begin{cases} I\cos(\pi/2)(1-2t) + J\sin(\pi/2)(1-2t), & t \in [0, 1/2]; \\ \gamma(2t-1), & t \in [1/2, 1]. \end{cases}$$

then we choose a symplectic path $\beta(t)$ in $\mathrm{Sp}(2n)^*_{L_0}$ starting from $\gamma(1)$ and ending at M_+ or M_-. We now define a joint path by

$$\bar{\gamma}(t) = \beta * \tilde{\gamma} := \begin{cases} \tilde{\gamma}(2t), & t \in [0, 1/2], \\ \beta(2t-1), & t \in [1/2, 1]. \end{cases}$$

By the definition, we see that the symplectic path $\bar{\gamma}$ starting from $-M_+$ and ending at either M_+ or M_-. As above, we define

$$\bar{\mathrm{Q}}(t) = \big(\bar{U}(t) - \sqrt{-1}\bar{V}(t)\big)\big(\bar{U}(t) + \sqrt{-1}\bar{V}(t)\big)^{-1}. \tag{A.4}$$

for $\bar{\gamma}(t) = \begin{pmatrix} \bar{S}(t) & \bar{V}(t) \\ \bar{T}(t) & \bar{U}(t) \end{pmatrix}$. We can choose a continuous function $\bar{\Delta}(t)$ in $[0, 1]$ such that

$$\det \bar{\mathrm{Q}}(t) = e^{2\sqrt{-1}\bar{\Delta}(t)}. \tag{A.5}$$

By the above arguments, we see that the number $\frac{1}{\pi}(\bar{\Delta}(1) - \bar{\Delta}(0)) \in \mathbb{Z}$ and it does not depend on the choice of the function $\bar{\Delta}(t)$.

Definition A.2. For a symplectic path $\gamma \in \mathrm{P}(2n)^*_{L_0}$, we define the L_0-index of γ by

$$i_{L_0}(\gamma) = \frac{1}{\pi}(\bar{\Delta}(1) - \bar{\Delta}(0)). \tag{A.6}$$

Definition A.3. For a symplectic path $\gamma \in \mathrm{P}(2n)^0_{L_0}$, we define the L_0-index of γ by

$$i_{L_0}(\gamma) = \inf\left\{ i_{L_0}(\tilde{\gamma}) \mid \tilde{\gamma} \in \mathrm{P}(2n)^*_{L_0}, \text{ and } \tilde{\gamma} \text{ is sufficiently close to } \gamma \right\}.$$

We note that $\Lambda(n) = U(n)/O(n)$; this means that for any linear subspace $L \in \Lambda(n)$, there is an orthogonal symplectic matrix

$$P = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}$$

with $A \pm \sqrt{-1}B \in U(n)$ such that $PL_0 = L$. P is uniquely determined by L up to an orthogonal matrix $C \in O(n)$. It means that for any other choice P' satisfying above conditions, there exists a matrix $C \in O(n)$ such that

$$P' = P \begin{pmatrix} C & 0 \\ 0 & C \end{pmatrix}.$$

See [McDuff and Salamon 1998, Lemma 2.31]. We define the conjugated symplectic path $\gamma_c \in \mathrm{P}(2n)$ of γ by $\gamma_c(t) = P^{-1}\gamma(t)P$.

Definition A.4. We define the L-nullity of any symplectic path $\gamma \in \mathrm{P}(2n)$ by

$$\nu_L(\gamma) \equiv \dim \ker_L(\gamma(1)) := \dim \ker V_c(1) = n - \operatorname{rank} V_c(1),$$

The $n \times n$ matrix function $V_c(t)$ is defined in (A.1) with the symplectic path γ replaced by γ_c, that is,

$$\gamma_c(t) = \begin{pmatrix} S_c(t) & V_c(t) \\ T_c(t) & U_c(t) \end{pmatrix}.$$

Definition A.5. For a symplectic path $\gamma \in \mathrm{P}(2n)$, we define the L-index of γ by

$$i_L(\gamma) = i_{L_0}(\gamma_c).$$

Theorem A.6. *If $\gamma \in \mathrm{P}(2n)^0_L$, there is a family of paths $\gamma_s \in \mathrm{P}(2n)_L$ depend continuous on $s \in [-1, 1]$ such that $\gamma_0 = \gamma$, $\gamma_s \in \mathrm{P}(2n)^*_L$, $s \neq 0$ and*

$$i_L(\gamma_s) - i_L(\gamma_{-s}) = \nu_L(\gamma) \quad \textit{for all } s \in (0, 1],$$

and

$$i_L(\gamma) = i_L(\gamma_{-s}), \; s \in (0, 1].$$

For a symmetric matrix function $B : [0, 1] \to \mathrm{L}_s(2n)$, we consider the functional

$$f(z) = \int_0^1 \big(\tfrac{1}{2}(-J\dot{z}, z) - (B(t)z, z)\big)\, dt, \quad z \in W_L,$$

where $W_L = \{z = (x, y)^T \in W^{1,2}([0, 1], \mathbb{R}^{2n}) \mid z(0), z(1) \in L\} \subset L^2$. By the saddle point reduction methods (see [Amann 1979; Amann and Zehnder 1980; Long 1993; 2002; Liu 2007]), there exists a finite-dimensional subspace X of W_L with $\dim X = 2d + n$ and an injection map $X \to W_L$, such that the function $a(x) = f(u(x))$ is C^2 and we have:

Theorem A.7. *For any $L \in \Lambda(n)$,*

$$\begin{aligned} m^-(a) &= d + i_L(B) + n, \\ m^0(a) &= \nu_L(B), \\ m^+(a) &= d - i_L(B) - \nu_L(B), \end{aligned}$$

where $m^*(a)$ for $* = +, 0, -$ are respectively the positive, null and the negative Morse indices of the function $a(x)$ at the origin.

Theorem A.8. *For any symplectic path $\gamma \in \mathrm{P}(2n)$, there holds*

$$i_{L_0}(\gamma) = \sum_{j=1}^n E\left(\frac{\theta_j(1) - \theta_j(0)}{2\pi}\right),$$

where $E(a) = \max\{k \in \mathbb{Z} \mid k < a\}$ and $\lambda_j(t) = e^{\sqrt{-1}\theta_j(t)}$ are the eigenvalues of $\mathrm{Q}(t)$ defined in (A.3).

Acknowledgement

A part of this paper was done while the author visiting Chern Institute of Mathematics, and the author would like to thank the Institute for partial support. The author appreciates the referee for his/her useful revision suggestions and for correcting some mistakes.

References

[Amann 1979] H. Amann, “Saddle points and multiple solutions of differential equations”, *Math. Z.* **169**:2 (1979), 127–166. MR 80j:47078 Zbl 0414.47042

[Amann and Zehnder 1980] H. Amann and E. Zehnder, “Nontrivial solutions for a class of nonresonance problems and applications to nonlinear differential equations”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **7**:4 (1980), 539–603. MR 82b:47077 Zbl 0452.47077

[Arnold 1986] V. I. Arnold, “The first steps of symplectic topology”, *Uspekhi Mat. Nauk* **41**:6(252) (1986), 3–18, 229. MR 89d:58034 Zbl 0618.58021

[Bott 1982] R. Bott, “Lectures on Morse theory, old and new”, *Bull. Amer. Math. Soc.* (*N.S.*) **7**:2 (1982), 331–358. MR 84m:58026a Zbl 0505.58001

[Chang 1981] K. C. Chang, "Solutions of asymptotically linear operator equations via Morse theory", *Comm. Pure Appl. Math.* **34**:5 (1981), 693–712. MR 82m:58015 Zbl 0444.58008

[Chekanov 1996] Y. V. Chekanov, "Hofer's symplectic energy and Lagrangian intersections", pp. 296–306 in *Contact and symplectic geometry (Cambridge, 1994)*, edited by C. B. Thomas, Publ. Newton Inst. **8**, Cambridge Univ. Press, Cambridge, 1996. MR 98b:58058 Zbl 0867.58026

[Chekanov 1998] Y. V. Chekanov, "Lagrangian intersections, symplectic energy, and areas of holomorphic curves", *Duke Math. J.* **95**:1 (1998), 213–226. MR 99k:58034 Zbl 0977.53077

[Clarke 1978] F. H. Clarke, "Solution périodique des équations hamiltoniennes", *C. R. Acad. Sci. Paris* **287** (1978), 951–952. Zbl 0422.35005

[Clarke 1979] F. H. Clarke, "A classical variational principle for periodic Hamiltonian trajectories", *Proc. Amer. Math. Soc.* **76**:1 (1979), 186–188. MR 84h:34071 Zbl 0434.70022

[Clarke 1981] F. H. Clarke, "Periodic solutions to Hamiltonian inclusions", *J. Differential Equations* **40**:1 (1981), 1–6. MR 83a:58035 Zbl 0461.34030

[Clarke and Ekeland 1978] F. H. Clarke and I. Ekeland, "Solutions périodiques, de période donnée, des équations hamiltoniennes", *C. R. Acad. Sci. Paris Sér. A-B* **287**:15 (1978), A1013–A1015. MR 80b:49025 Zbl 0404.34033

[Clarke and Ekeland 1980] F. H. Clarke and I. Ekeland, "Hamiltonian trajectories having prescribed minimal period", *Comm. Pure Appl. Math.* **33**:2 (1980), 103–116. MR 81e:70017 Zbl 0403.70016

[Clarke and Ekeland 1982] F. H. Clarke and I. Ekeland, "Nonlinear oscillations and boundary value problems for Hamiltonian systems", *Arch. Rat. Mech. Anal.* **78**:4 (1982), 315–333. MR 83h:58038 Zbl 0514.34032

[Conley and Zehnder 1984] C. Conley and E. Zehnder, "Morse-type index theory for flows and periodic solutions for Hamiltonian equations", *Comm. Pure Appl. Math.* **37**:2 (1984), 207–253. MR 86b:58021 Zbl 0559.58019

[Dong 2005] Y. Dong, "Maslov type index theory for linear Hamiltonian systems with Bolza boundary value conditions and multiple solutions for nonlinear Hamiltonian systems", *Pacific J. Math.* **221**:2 (2005), 253–280. MR 2006k:37163 Zbl 05031482

[Ekeland 1990] I. Ekeland, *Convexity methods in Hamiltonian mechanics*, vol. 19, Ergebnisse der Mathematik, Springer, Berlin, 1990. MR 91f:58027 Zbl 0707.70003

[Fei and Qiu 1996] G. Fei and Q. Qiu, "Minimal period solutions of nonlinear Hamiltonian systems", *Nonlinear Anal.* **27**:7 (1996), 821–839. MR 97i:58143 Zbl 0858.34034

[Floer 1988] A. Floer, "Morse theory for Lagrangian intersections", *J. Differential Geom.* **28**:3 (1988), 513–547. MR 90f:58058 Zbl 0674.57027

[Floer 1989] A. Floer, "Cuplength estimates on Lagrangian intersections", *Comm. Pure Appl. Math.* **42**:4 (1989), 335–356. MR 90g:58034 Zbl 0683.58017

[Girardi and Matzeu 1991] M. Girardi and M. Matzeu, "Dual Morse index estimates for periodic solutions of Hamiltonian systems in some nonconvex superquadratic cases", *Nonlinear Anal.* **17**:5 (1991), 481–497. MR 92k:58224 Zbl 0779.34033

[Guo and Liu 2007] F. Guo and C. Liu, "Multiplicity of Lagrangian orbits on symmetric star-shaped hypersurfaces", *Nonlinear Anal.* (2007).

[Hofer 1988] H. Hofer, "Lusternik-Schnirelman-theory for Lagrangian intersections", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **5**:5 (1988), 465–499. MR 91b:58068 Zbl 0669.58006

[Liu 2001] C. G. Liu, "Dual Morse index for subquadratic Hamiltonian systems and the stability of the closed characteristics", *Acta Math. Sinica* (*Chin. Ser.*) **44** (2001), 1073–1088. MR 2002j:37090 Zbl 1012.37047

[Liu 2005a] C.-G. Liu, "Cup-length estimate for Lagrangian intersections", *J. Differential Equations* **209**:1 (2005), 57–76. MR 2005g:53176 Zbl 02136639

[Liu 2005b] C.-G. Liu, "A note on the monotonicity of the Maslov-type index of linear Hamiltonian systems with applications", *Proc. Roy. Soc. Edinburgh Sect. A* **135**:6 (2005), 1263–1277. MR 2007d:53135 Zbl 05001218

[Liu 2007] C. Liu, "Maslov-type index theory for symplectic paths with Lagrangian boundary conditions", *Advanced Nonlinear Studies* **7**:1 (2007), 131–161.

[Long 1993] Y. Long, *Index Theory of Hamiltonian Systems with Applications*, Science Press, Beijing, 1993. In Chinese.

[Long 2002] Y. Long, *Index theory for symplectic paths with applications*, Progress in Mathematics **207**, Birkhäuser Verlag, Basel, 2002. MR 2003d:37091 Zbl 1099.81543

[McDuff and Salamon 1998] D. McDuff and D. Salamon, *Introduction to symplectic topology*, Second ed., Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 1998. MR 2000g:53098 Zbl 1066.53137

[Mohnke 2001] K. Mohnke, "Holomorphic disks and the chord conjecture", *Ann. of Math.* (2) **154**:1 (2001), 219–222. MR 2002f:53147 Zbl 1007.53065

[Oh 1995] Y.-G. Oh, "Floer cohomology of Lagrangian intersections and pseudo-holomorphic disks. III. Arnold-Givental conjecture", pp. 555–573 in *The Floer memorial volume*, edited by H. Hofer et al., Progr. Math. **133**, Birkhäuser, Basel, 1995. MR 96k:58037 Zbl 0842.58033

[Ono 1996] K. Ono, "Lagrangian intersection under Legendrian deformations", *Duke Math. J.* **85**:1 (1996), 209–225. MR 97g:58036 Zbl 0868.58035

Received June 6, 2006. Revised October 14, 2006.

CHUN-GEN LIU
SCHOOL OF MATHEMATICS AND LPMC
NANKAI UNIVERSITY
TIANJIN 300071
CHINA

liucg@nankai.edu.cn

Guidelines for Authors

Authors may submit manuscripts at pjm.math.berkeley.edu/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 Web site in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 232 No. 1 September 2007

0030-8730(200709)232:1;1-C