Pacific Journal of Mathematics

Volume 264 No. 2 August 2013

PACIFIC JOURNAL OF MATHEMATICS

msp.org/pjm

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

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

The subscription price for 2013 is US $400/year for the electronic version, and $485/year for print and electronic.
Subscriptions, requests for back issues and changes of subscribers address should be sent to Pacific Journal of Mathematics, P.O. Box 4163, Berkeley, CA 94704-0163, U.S.A. The Pacific Journal of Mathematics is indexed by Mathematical Reviews, Zentralblatt MATH, PASCAL CNRS Index, Referativnyi Zhurnal, Current Mathematical Publications and the Science Citation Index.

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

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

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

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.257

ON 4-MANIFOLDS, FOLDS AND CUSPS

STEFAN BEHRENS

We study *simple wrinkled fibrations*, a variation of the simplified purely wrinkled fibrations of Williams (*Geom. Topol.* 14:2 (2010), 1015–1061), and their combinatorial description in terms of *surface diagrams*. We show that simple wrinkled fibrations induce handle decompositions of their total spaces which are very similar to those obtained from Lefschetz fibrations. The handle decompositions turn out to be closely related to surface diagrams and we use this relationship to interpret some well known operations on 4-manifolds in terms of surface diagrams. This, in turn, allows us classify all closed 4-manifolds which admit simple wrinkled fibrations of genus one, the lowest possible fiber genus.

1. Introduction

After the pioneering work of Donaldson [1999] and Gompf [1999] on symplectic 4-manifolds and Lefschetz fibrations and of Auroux, Donaldson and Katzarkov on near-symplectic 4-manifolds [Auroux et al. 2005], the study of singular fibration structures on smooth 4-manifolds has received considerable attention in the research literature. Among the highlights in the field have been existence results for so called *broken Lefschetz fibrations* over the 2-sphere on all closed, oriented 4-manifolds [Akbulut and Karakurt 2008; Baykur 2008; Gay and Kirby 2007; Lekili 2009] as well as a classification of these maps up to homotopy [Lekili 2009; Williams 2010]. Furthermore, the classical observation that Lefschetz fibrations over the 2-sphere are accessible via handlebody theory and can be described more or less combinatorially in terms of collections of simple closed curves on a regular fiber known as the *vanishing cycles* [Kas 1980; Gompf and Stipsicz 1999] was extended to the broken Lefschetz setting in [Baykur 2009].

Our starting point is the work of Williams [2010], who introduced the closely related notion of *simplified purely wrinkled fibrations*, proved their existence and exhibited a similar combinatorial description of these maps — again by collections of simple closed curves on a regular fiber — which he calls *surface diagrams*. It

MSC2010: primary 57M50; secondary 57R65.
Keywords: 4-manifolds, folds, cusps, simple wrinkled fibrations, simplified purely wrinkled fibrations, broken Lefschetz fibrations, surface diagram.

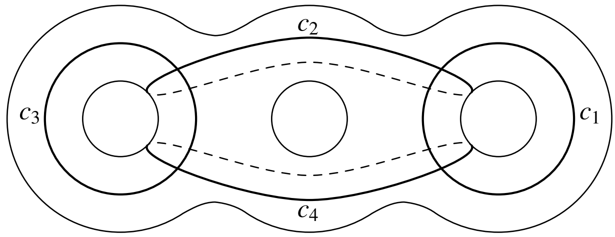

Figure 1. A surface diagram of $S^1 \times S^3 \# S^1 \times S^3$ due to Hayano [2012].

follows that all smooth, closed, oriented 4-manifolds can be described by surface diagrams; an example of such a diagram is shown in Figure 1. However, the correspondence between simplified purely wrinkled fibrations and surface diagrams has been somewhat unsatisfactory in that it usually involved arguments using broken Lefschetz fibrations and the assumption that the fiber genus is sufficiently high.

It is one of our goals to provide a detailed and intrinsic account of this correspondence and to clarify the situation in the lower-genus cases. Once this is done we give some applications.

We now describe the contents of this paper in more detail. In Section 2 we begin by recalling some preliminaries from the singularity theory of smooth maps and the theory of mapping class groups of surfaces. This section is slightly lengthy because we intend to use it as a reference for future work.

The following two sections form the technical core of this paper. In Section 3 we introduce *simple wrinkled fibrations* over a general base surface; in the case when the base is the 2-sphere our definition is almost equivalent to Williams' simplified purely wrinkled fibrations and our reason for introducing a new name is mainly to reduce the number of syllables. We explain how the study of simple wrinkled fibrations reduces to certain fibrations over the annulus which we call *annular simple wrinkled fibrations* to which we associate *twisted surface diagrams*; roughly, such a diagram consists of a closed, oriented surface Σ, an ordered collection of simple closed curves $c_1, \ldots, c_l \subset \Sigma$ and an orientation-preserving diffeomorphism $\mu : \Sigma \to \Sigma$ such that pairs of consecutive curves (c_i and c_{i+1} for $i < l$, as well as $\mu(c_l)$ and c_1) intersect transversely in one point. We prove the following:

Theorem 1.1. *There is a bijective correspondence between annular simple wrinkled fibrations up to equivalence and twisted surface diagrams up to equivalence.*

For precise definitions we refer to Section 3. In the course of the proof we show that annular simple wrinkled fibrations induce (relative) handle decompositions of their total spaces which are, in fact, encoded in a twisted surface diagram (Section 3B). These handle decompositions bear a very close resemblance with those obtained from Lefschetz fibrations; the only difference appears in the framings of certain 2-handles. The section ends with an investigation of the ambiguities for gluing surface bundles to the boundary components of annular simple wrinkled fibrations.

In Section 4 we specialize to the case when the base surface is either a disk or a sphere and recover Williams' setting. Using our results about annular simple wrinkled fibrations we obtain a precise correspondence between Williams' (untwisted) surface diagrams and simple wrinkled fibrations over the disk (Proposition 4.1) and the sphere (Corollary 4.2). In particular, our approach provides a direct way to construct a simple wrinkled fibration from a given surface diagram circumventing the previously necessary detour via broken Lefschetz fibrations.[1]

Next, we address the subtle question of which surface diagrams give rise to simple wrinkled fibrations over the sphere and thus describe closed 4-manifolds. Just as in the theory of Lefschetz fibrations, the key is to understand the boundary of the associated simple wrinkled fibration over the disk. We show how to identify this boundary with a mapping torus and describe its monodromy in terms of the surface diagram. Unfortunately, it turns out that the boundary is much harder to understand than in the Lefschetz setting.

We then go on to review the handle decompositions exhibited in Section 3 when the base is the disk or the sphere and describe a recipe for drawing Kirby diagrams for them. To complete the picture, we compare our decompositions with the ones obtained via simplified broken Lefschetz fibrations.

In Sections 5 and 6 we give some applications. We show that certain substitutions of curve configurations in surface diagrams correspond to cut-and-paste operations on 4-manifolds. In particular, we give a surface diagram interpretation of blow-ups and sum stabilizations, by which we mean connected sums with $\mathbb{C}P^2$, $\overline{\mathbb{C}P^2}$ and $S^2 \times S^2$. Using these we easily obtain a classification of closed 4-manifolds which admit simple wrinkled fibrations with the lowest possible fiber genus.

Theorem 1.2. *A smooth, closed, oriented 4-manifold admits a simple wrinkled fibration of genus one if and only if it is diffeomorphic to $kS^2 \times S^2$ or $m\mathbb{C}P^2 \# n\overline{\mathbb{C}P^2}$ where $k, m, n \geq 1$.*

Our result should be compared to [Baykur and Kamada 2010] and [Hayano 2011], where the classification problem of genus-one simplified broken Lefschetz fibrations is addressed but only partial solutions are achieved. However, it should also be noted that their class of maps is strictly larger than that of genus-one simple wrinkled fibrations and it is thus conceivable that the classification is more complicated.

Section 7 closes this paper by highlighting what we consider as some of the main problems in the field and by outlining some related developments.

[1] By now this can be considered as a special case of [Gay and Kirby 2012], which appeared while we were writing this paper.

Conventions. By default, all manifolds are smooth, compact and orientable; all maps are smooth and all diffeomorphisms preserve orientations. Given a submanifold $S \subset M$ we denote by νS (respectively $\bar{\nu}S$) an open (respectively closed) tubular neighborhood of S and whenever we speak of neighborhoods of submanifolds we usually mean tubular neighborhoods. For induced orientations on boundaries we use the *outward normal first* convention and, in order to coherently orient regular fibers of maps between oriented manifolds, we use the *fiber first* convention. Exceptions to these rules will be explicitly stated and we reserve the right to sometimes restate some of the conditions for emphasis.

2. Preliminaries

To fix some terminology, let $f : M \to N$ be a smooth map with differential $df : TM \to TN$. A *critical point* or *singularity* of f is a point $p \in M$ such that df_p is not surjective. The set of critical points, called the *critical locus* of f, will be denoted by

$$\mathcal{C}_f := \{p \in M \mid \operatorname{rk} df_p < \dim N\} \subset M.$$

The image of a critical point is called a *critical value* and the set of all critical values is called the *critical image* of f.

As customary, we call the preimage of a point a *fiber*, usually decorated with the adjectives regular or singular indicating whether or not the fiber contains singularities. Note that regular fibers are always smooth submanifolds with trivial normal bundle.

Remark 2.1. The terms critical point and singularity are used synonymously and somewhat inconsistently in the literature, even in standard references such as [Golubitsky and Guillemin 1973]. We will adapt to this custom of arbitrariness and also use both terms depending on which seems more appropriate. However, we would like to stress that neither term indicates the failure of a map to be smooth at a given point — all maps we consider are smooth — they just indicate irregular behavior of the differential at that point as described above.

2A. ***Folds, cusps and Lefschetz singularities.*** As a warm-up, recall that a generic map from any compact manifold to a 1-dimensional manifold has only finitely many critical points on which it is injective and, moreover, all critical points are of *Morse type*; that is, they are locally modeled on maps of the form

$$(x_1, \dots, x_n) \mapsto -x_1^2 - \dots - x_k^2 + x_{k+1}^2 + \dots + x_n^2,$$

where the number k is called the (*Morse*) *index* of the critical point. (We say that a map $f : M^m \to N^n$ is *locally modeled around* $p \in M$ on $f_0 : \mathbb{R}^m \to \mathbb{R}^n$ if there are local coordinates around p and $f(p)$ mapping these points to the origin such that the coordinate representation of f agrees with f_0.) Maps whose critical points are all of Morse type are called *Morse functions*.

A similar statement holds for maps to surfaces. For convenience we take the source to be 4-dimensional from now on. In this setting the Morse critical points are replaced by two other types of singularities known as *folds* and *cusps* which can also be described in terms of local models. The model for a fold point is the map $\mathbb{R}^4 \to \mathbb{R}^2$ given by the formula

$$(t, x, y, z) \mapsto (t, -x^2 - y^2 \pm z^2) \tag{2-1}$$

and the cusps are locally modeled on

$$(t, x, y, z) \mapsto (t, -x^3 + 3tx - y^2 \pm z^2). \tag{2-2}$$

If the sign in either of the above expressions is positive (respectively negative), then the singularity is called *indefinite* (respectively *definite*).

An easy calculation shows that the critical loci of the fold and cusp models are given by $\{ (r, 0, 0, 0) \mid r \in \mathbb{R} \}$ and $\{ (r^2, r, 0, 0) \mid r \in \mathbb{R} \}$, respectively. As a consequence, the critical image of a smooth map is a smooth 1-dimensional submanifold near fold and cusp points. The critical images of both models are shown in Figure 2. Note that the critical image is smoothly embedded in the fold model whereas in the cusp case it is topologically embedded via a smooth homeomorphism whose inverse fails to be smooth only at the cusp point.

It follows directly from the models that folds always come in 1-dimensional families on which the map restricts to an immersion. We will usually be sloppy and refer to such an arc of fold points in the source as well as their image in the target as *fold arcs*. Furthermore, cusps are isolated in the critical locus in the sense that there is a small neighborhood which contains no other cusps. However, cusps are not isolated singularities. In fact, one can show that any cusp is surrounded by two fold arcs, at least one of which is indefinite.

We can now state the normal form of generic maps from 4-manifolds to surfaces.

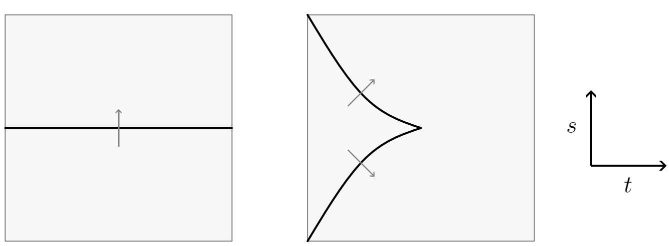

Figure 2. The critical images of the fold and cusp models.

Theorem 2.2 (generic maps to surfaces). *A generic smooth map from a* 4*-manifold to a surface has only fold and cusp singularities*, *it is injective on the cusps and restricts to an immersion of its critical locus with only transverse intersections between fold arcs.*

Results of this kind are common knowledge in singularity theory; precise references for Theorem 2.2 are [Golubitsky and Guillemin 1973, Theorem 5.2] and [Levine 1964, Theorem 1] (see also [Boardman 1967; Morin 1965]).

The preceding discussion shows, in particular, that the critical locus of a generic map to a surface is a smooth 1-dimensional submanifold of the source.

Remark 2.3. Recently, these generic maps to surfaces have appeared under the name *Morse 2-functions* in [Gay and Kirby 2011a; 2011b; 2012].

In what follows we only deal with indefinite singularities. So from now on, when we speak of folds and cusps, we always mean the indefinite ones.

Figure 2 contains some further decorations which we will now explain. Both folds and cusps are intimately related to 3-dimensional Morse–Cerf theory. The fold models a trivial homotopy of a Morse functions with one critical point (of index two) on the vertical slices. This means that the model restricted to a small arc transverse to the fold locus is a Morse function with one critical point of index one or two, depending on the direction. The arrows in the picture indicate the direction in which the index is two. Note that the topology of the fibers of either side of a fold arc is necessarily different.

Similarly, the cusp is also a homotopy of Morse functions on the vertical slices, although a nontrivial one. It models the cancellation of a pair of critical points of index one and two. The arrows indicate the index two direction of the fold arcs adjacent to the cusp.

For the moment, this is all we have to say about folds and cusps. Another important type of singularity which has its roots in (complex) algebraic geometry is the *Lefschetz singularity* and its local model is given in complex coordinates by

$$L : \mathbb{C}^2 \to \mathbb{C}, \quad (z, w) \mapsto zw.$$

At this point it becomes important whether the charts that we use to model the map are orientation-preserving. Indeed, the use of orientation-reversing charts for the Lefschetz model produces so called *achiral* Lefschetz singularities which are not compatible with complex geometry; in orientation-preserving coordinates achiral Lefschetz singularities can be modeled by $(z, w) \mapsto \bar{z}w$ which is not holomorphic. We will thus always use orientation-preserving charts to model singularities whenever the source or target are oriented. Note that this is no restriction for folds and cusps since both models admit an orientation-reversing diffeomorphism which leaves the map invariant.

As stated in the introduction, maps with (indefinite) fold, cusp and Lefschetz singularities have been prominently featured in the research literature over the past decade. Unfortunately, different authors have used different names for various types of maps and there is yet no commonly accepted terminology in the field. For the purpose of this paper we use this:

Definition 2.4. Let $f : X \to B$ be a surjective map from an oriented 4-manifold to an oriented surface, with critical locus $\mathcal{C}_f$. Assume that all intersections in the critical image are transverse intersections of fold arcs and $\mathcal{C}_f$ is transverse to the boundary of X. We call

(a) a *wrinkled fibration* if $\mathcal{C}_f$ contains only indefinite folds and cusps,

(b) a (*broken*) *Lefschetz fibration* if $\mathcal{C}_f$ contains only Lefschetz singularities (and indefinite folds),

(c) a *broken fibration* if $\mathcal{C}_f$ contains only indefinite folds, cusps and Lefschetz singularities.

We will usually refer to X as the *total space* and to B as the *base* of f.

If $f : X \to B$ is a broken fibration, then $\partial X \cap \mathcal{C}_f$ is either empty or consists of finitely many fold points and it follows from the fold model that f restricts to a circle valued Morse function over each boundary component of B.

The regular fibers of f are orientable surfaces and our conventions determine an orientation. We will usually assume that $\partial X = f^{-1}(\partial B)$ so that the fibers are closed surfaces.

It is quite useful to think of broken fibrations as singular families of surfaces parametrized by the base. More precisely, the images of the folds and cusps cut the base into several regions which may or may not contain Lefschetz singularities. Each regular fiber is an orientable surface whose topological type depends only on the region that it maps into. One thus decorates the base with the topological type of the fibers over each region together with some information about what happens to a fiber if one crosses a fold arc (the fold vanishing cycles corresponding to the little arrows we have indicated above, see Definition 3.11) or runs into a Lefschetz singularity (the Lefschetz vanishing cycle). Under certain circumstances this data is enough to determine the map as we will see later on; see also [Gay and Kirby 2012].

We finish this section with a short review of the homotopy classification of broken fibrations over S^2 that was mentioned in the introduction. An important contribution of Lekili [2009] is that he showed how to pass back and forth between broken Lefschetz fibrations and wrinkled fibrations via two *local homotopies*, i.e., homotopies supported in arbitrarily small balls. As portrayed in Figure 3 one can *wrinkle* a Lefschetz point into an indefinite triangle (that is, an indefinite circle with

Figure 3. Wrinkling (left) and unsinking (right) a Lefschetz singularity.

three cusps) and one can exchange a cusp for a Lefschetz singularity; this move is sometimes called *unsinking* a Lefschetz point from a fold. (Moreover, he showed that these modifications work equally well with achiral Lefschetz singularities which, together with the results of [Gay and Kirby 2007], proves the existence of broken Lefschetz fibrations.) As a consequence, one can translate questions about broken fibrations into questions about wrinkled fibrations which are accessible by means of singularity theory. For example, there is a structural result similar to Theorem 2.2 for generic homotopies between wrinkled fibrations. The basic building blocks include isotopies of the base and total space and three types of modifications (and their inverses) that are realized by local homotopies: the *birth/death*, the *merge* and the *flip*. Figure 4 shows their effect on the critical image. In general, such a generic homotopy will pass through maps with definite singularities. However, the main theorem in [Williams 2010], which was conjectured in [Lekili 2009], states that indefinite singularities can, in fact, be avoided. In other words, any two homotopic wrinkled fibrations are homotopic through wrinkled fibrations.

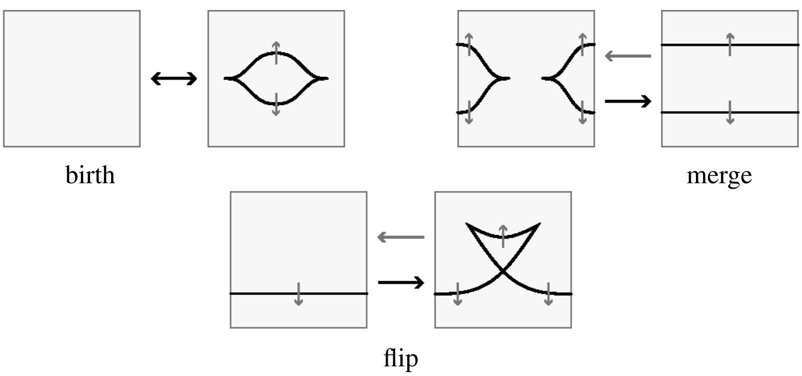

Figure 4. The basic local homotopies.

Remark 2.5. It has become common to refer to an application of any of these modifications as *moves* performed on a broken fibration; this terminology is due to [Lekili 2009]. It is important to note that most of these moves are not strictly reversible in the following sense. If the critical image of a given broken fibration exhibits the left configuration in any of the pairs, it is always possible to replace it by the one on the right. However, it might not be possible to go into the other direction. The only exception is the birth. In all other cases some extra conditions are needed to go from right to left. This is indicated in our pictures with shaded arrows. For further details we refer to [Lekili 2009].

Remark 2.6. There is some disagreement in the literature about which direction in the second pair in Figure 4 should be called merge and which inverse merge. To avoid this decision we simply speak of *merging cusps* and *merging folds*, respectively.

2B. ***Surfaces and simple closed curves.*** As we pointed out, the regular fibers of broken fibrations are surfaces and these fibers will be our main focus later on. The theory of surfaces and mapping class groups is yet another field of mathematics with many different conventions and, in the author's experience, it can be confusing to decide whether a statement in some reference actually applies to a situation at hand. For this reason we give very precise definitions, deliberately risking to be overly precise.

By a *surface* Σ we mean a compact, orientable, 2-dimensional manifold, possibly with boundary and some marked points in the interior. A *simple closed curve* in Σ is a closed, connected, 1-dimensional submanifold of Σ that does not meet the boundary or the marked points. We usually consider simple closed curves up to ambient isotopy in Σ relative to $\partial\Sigma$ and the marked points and will not make a notational distinction between a simple closed curve and its isotopy class. Note that according our definition simple closed curves are unoriented objects. However, from time to time it will be convenient to choose orientations on them in order to speak of their homology classes.

The *geometric intersection number* of two simple closed curves $a, b \subset \Sigma$ is

$$i(a,b) := \min\{\#(\alpha\cap\beta) \mid \alpha\sim a,\ \beta\sim b,\ \alpha\pitchfork\beta\} \in \mathbb{N},$$

where the signs $\sim$ and $\pitchfork$ indicate isotopy and transverse intersection. If the curves as well as the surface are oriented, then we also have an *algebraic intersection number* which is obtained by a signed count of intersections after making the curves transverse. Equivalently, this number can be described as

$$\langle a,b\rangle := \langle [a],[b]\rangle_\Sigma := \langle [a],[b]\rangle_{H_1(\Sigma)} \in \mathbb{Z},$$

where the bracket on the right side denotes the intersection form on $H_1(\Sigma)$. (In the present paper homology is always taken with integer coefficients.)

Note that the algebraic intersection number is alternating and depends only on the homology classes of the oriented simple closed curves while the geometric intersection number is symmetric and depends on the isotopy classes. Both intersection numbers have the same parity and satisfy the inequality

$$|\langle a,b\rangle| \le i(a,b). \tag{2-3}$$

We say that a and b are *geometrically dual* (respectively *algebraically dual*) if their geometric (respectively algebraic) intersection number is one.

A simple closed curve $a \subset \Sigma$ is called *nonseparating* if its complement is connected, otherwise it is called *separating*. Note that a simple closed curve is separating if and only if it is null-homologous (with either orientation) and thus simple closed curves that have geometric or algebraic duals are automatically nonseparating.

Diffeomorphisms of surfaces. Let us now turn to diffeomorphisms of surfaces. Let $\mathrm{Diff}^+(\Sigma, \partial\Sigma)$ denote the set of orientation-preserving diffeomorphisms that restrict to the identity on $\partial\Sigma$ and preserve the set of marked points. The *mapping class group* of Σ is defined as

$$\mathrm{Mod}(\Sigma) := \pi_0(\mathrm{Diff}^+(\Sigma, \partial\Sigma), \mathrm{id}).$$

Given a simple closed curve $a \subset \Sigma$ there is a well defined mapping class $\tau_a \in \mathrm{Mod}(\Sigma)$ called the (right-handed) *Dehn twist* about a. Similarly, any simple arc $r \subset \Sigma$ that connects two distinct marked points gives rise to a *half twist* $\bar{\tau}_r \in \mathrm{Mod}(\Sigma)$.

It is well known that $\mathrm{Mod}(\Sigma)$ is generated by the collection of Dehn twist and half twists, where the latter are only needed in the presence of marked points. On the other hand, mapping classes can be effectively studied by their action on (isotopy classes of) simple closed curves. In particular, it is desirable to understand the effect of Dehn twists on simple closed curves. While this can be tricky, the situation simplifies significantly on the level of homology classes.

Proposition 2.7 (Picard–Lefschetz formula). *Let Σ be a surface, $a \subset \Sigma$ a simple closed curve and let $x \in H_1(\Sigma)$. Then for any orientation on a we have*

$$(\tau_a^k)_* x = x + k\,\langle [a], x\rangle\,[a]. \tag{2-4}$$

In particular, if b is an oriented simple closed curve, then

$$[\tau_a^k(b)] = [b] + k\,\langle [a], [b]\rangle\,[a]. \tag{2-5}$$

Proof. See [Farb and Margalit 2011, Proposition 6.3] □

Remark 2.8. The Picard–Lefschetz formula is particularly useful for the torus since, in that case, mapping classes are completely determined by their action on homology.

Another useful tool is the *change of coordinates principle*, which roughly states that any two configurations of simple closed curves on a surface with the same intersection pattern can be mapped onto each other by a diffeomorphism. We will only use the following special cases. For details we refer to [Farb and Margalit 2011, Chapter 1.3].

Proposition 2.9 (change of coordinates principle). *If $a, b \subset \Sigma$ is a pair of nonseparating simple closed curves, then there exists some $\phi \in \mathrm{Diff}^+(\Sigma, \partial\Sigma)$ such that $\phi(a) = b$. Furthermore, if a, b and a', b' are two pairs of geometrically dual curves, then there is some $\phi \in \mathrm{Diff}^+(\Sigma, \partial\Sigma)$ such that $\phi(a) = a'$ and $\phi(b) = b'$.*

Mapping tori and their automorphisms. Given a surface Σ and a diffeomorphism $\mu : \Sigma \to \Sigma$, possibly not orientable or orientation-preserving, we can form the *mapping torus*

$$\Sigma(\mu) := \big(\Sigma \times [0, 1]\big)/\big((x, 1) \sim (\mu(x), 0)\big)$$

which is a 3-manifold that fibers over $S^1 \cong [0, 1]/\{0, 1\}$ in the obvious way. If Σ is oriented and μ is orientation-preserving, then our conventions stated in the introduction induce an orientation on $\Sigma(\mu)$. All surface bundles over S^1 can be described as mapping tori. Indeed, if a 3-manifold fibers over S^1, then one chooses a fiber and a lift of a vector field that determines the orientation of S^1 and the return map of the flow of this vector field induces a diffeomorphism of the fiber which is usually called the *monodromy*.

Let Y be an oriented 3-manifold that fibers over the circle via a map $f : Y \to S^1$. An *automorphism* of (Y, f) is an orientation- and fiber-preserving diffeomorphism of Y. We denote the group of automorphisms by $\mathrm{Aut}(Y, f)$ or simply by $\mathrm{Aut}(Y)$ when the fibration is clear from the context. If we identify Y with a mapping torus, say $\Sigma(\mu)$, then we obtain a description of $\mathrm{Aut}(Y)$ in terms of diffeomorphisms of Σ. Indeed, any element $\phi \in \mathrm{Aut}(\Sigma(\mu))$ can be considered as a path $(\phi_t)_{t\in[0,1]}$ in $\mathrm{Diff}^+(\Sigma)$ connecting some element $\phi_0 \in \mathrm{Diff}^+(\Sigma)$ to $\phi_1 = \mu^{-1}\phi_0\mu$. In particular, ϕ_0 must be isotopic to $\mu^{-1}\phi_0\mu$ and thus represents an element of $C_{\mathrm{Mod}(\Sigma)}(\mu)$, the centralizer in $\mathrm{Mod}(\Sigma)$ of (the mapping class represented by) μ. Elaborating on this observation one arrives at the conclusion that

$$\pi_0\big(\mathrm{Aut}(Y)\big) \cong \pi_0\big(\mathrm{Aut}(\Sigma(\mu))\big) \cong C_{\mathrm{Mod}(\Sigma)}(\mu) \ltimes \pi_1(\mathrm{Diff}(\Sigma), \mathrm{id}), \tag{2-6}$$

where the multiplication on the right side is given by

$$(g, \sigma) \cdot (h, \tau) = \big(h \circ g, (g^{-1}\tau g) * \sigma\big).$$

This means that there are essentially two types of automorphism of mapping tori: the ones that are constant on the fibers coming from $C_{\mathrm{Mod}(\Sigma)}(\mu)$ and the ones coming from $\pi_1(\mathrm{Diff}(\Sigma), \mathrm{id})$ that vary with the fibers and restrict to the identity on the reference fiber. However, it turns out that for most surfaces there are no nonconstant automorphisms.

Theorem 2.10 [Earle and Eells 1969]. *If Σ is a closed, orientable surface without marked points, the group $\pi_1(\mathrm{Diff}(\Sigma), \mathrm{id})$ is isomorphic to $\mathbb{Z}_2$, $\mathbb{Z} \oplus \mathbb{Z}$, or the trivial group, depending on whether the genus g equals* 0, 1, *or more than* 1.

Hence, as soon as the genus of the fiber of a mapping torus is at least two, all automorphisms are isotopic (through automorphisms) to constant ones.

Remark 2.11. It is important not to confuse the group $\mathrm{Aut}(Y)$ with the group of all (orientation-preserving) diffeomorphisms of Y. A general diffeomorphism will not even be isotopic to a fiber-preserving one!

Theorem 2.10 has many important consequences, of which we only highlight one.

Corollary 2.12. *Let $P \to S^2$ be a surface bundle with closed fibers of genus g.*

(1) *If $g = 0$, then P is diffeomorphic to $S^2 \times S^2$ or $\mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$.*

(2) *If $g = 1$, then P is diffeomorphic to $T^2 \times S^2$, $S^1 \times S^3$ or $S^1 \times L(n, 1)$.*

(3) *If $g \geq 2$, then P is diffeomorphic to $\Sigma_g \times S^2$*

Proof. For the genus-one case see [Baykur and Kamada 2010, Lemma 10]. The other cases are well known. □

3. Simple wrinkled fibrations over general base surfaces

Without further ado we introduce the main objects of study in this paper.

Definition 3.1. Let X be a 4-manifold and B a surface, both oriented. A *simple wrinkled fibration* with *total space* X and *base* B is wrinkled fibration $w : X \to B$ with the following additional properties:

(1) $\partial X = w^{-1}(B)$.

(2) $\mathcal{C}_w \cap \partial X = \varnothing$.

(3) $\mathcal{C}_w$ is nonempty, connected, and contains a cusp.

(4) w is injective on $\mathcal{C}_w$.

(5) All fibers of w are connected.

The *genus* of w is the maximal genus among all regular fibers. Finally, two simple wrinkled fibrations $w : X \to B$ and $w' : X' \to B'$ are *equivalent* if there are orientation-preserving diffeomorphisms $\hat{\phi} : X \to X'$ and $\check{\phi} : B \to B'$ such that $w' \circ \hat{\phi} = \check{\phi} \circ w$.

A neighborhood of the critical image of a simple wrinkled fibration is shown in Figure 5. Before we continue we make some remarks about the definition.

Remark 3.2. Simple wrinkled fibrations over S^2 are, in essence, the same as Williams' simplified purely wrinkled fibrations, with two minor differences. On the one hand we do not put restrictions on the fiber genus, but on the other we require the presence of cusps. Both conditions can always be achieved by applying

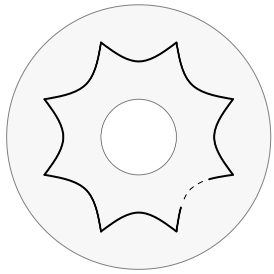

Figure 5. A neighborhood of the critical image of a simple wrinkled fibration.

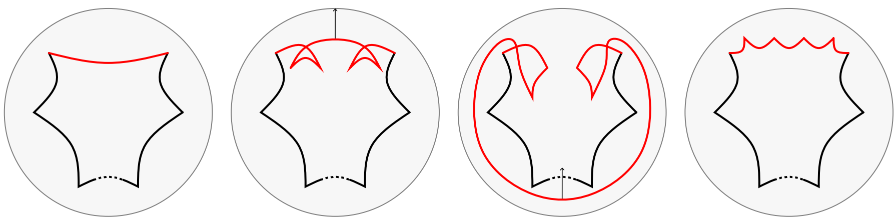

Figure 6. The base diagrams during a flip-and-slip move. (The pictures show the complement of a disk in the lower-genus region of the original fibration.)

a *flip-and-slip move* (see next remark) and are thus not restrictive. Moreover, the "simple wrinkled fibrations without cusps" are easily classified (see Example 3.7), so one does not lose too much by excluding them.

Remark 3.3. Given a simple wrinkled fibration over S^2 there is an important homotopy to another such simple wrinkled fibration which has become known as a *flip-and-slip move*. Its effect on the base diagram is shown in Figure 6. One first perform two flips on the same fold arc and then chooses an isotopy of the total space (the *slip*) during which the critical image undergoes the changes demonstrated in the picture. A flip-and-slip increases the fiber genus by one and introduces four new cusps.

Remark 3.4. In spite of the lengthy definition, simple wrinkled fibrations are arguably the simplest possible maps from 4-manifolds to surfaces, at least as far as their singularity structure is concerned. As will be explained in detail it is this simplicity which makes it possible to give nice combinatorial descriptions of 4-manifolds.

Given the rather specialized nature of simple wrinkled fibrations one might wonder whether they actually exist. This is indeed the case and we begin by giving some simple constructions.

Example 3.5 (surface bundles). Let $\pi : X \to B$ be a surface bundle over a surface B with closed fibers of genus g. Then we can perform a birth homotopy on π to obtain a simple wrinkled fibration of genus $g+1$ with two cusps.

Example 3.6 (Lefschetz fibrations). If $f : X \to B$ is a Lefschetz fibration (possibly achiral) with closed fibers of genus g, then after wrinkling all the Lefschetz singularities we obtain a number of disjoint circles with three cusps in the critical image. By suitably merging cusps we can turn this configuration into a single circle resulting in a simple wrinkled fibration of genus $g + 1$.

Example 3.7 (the case without cusps). This example includes the broken Lefschetz fibration on S^4 from [Auroux et al. 2005] that was mentioned in the introduction. Let Ω be a cobordism from Σ_g to Σ_{g-1} together with a Morse function $\mu : \Omega \to I$ with exactly one critical point of index two. Then $\mu \times \mathrm{id} : \Omega \times S^1 \to I \times S^1$ is a stable map with one circle of indefinite folds which fails to be a simple wrinkled fibration only because it does not have any cusps. Nevertheless, we can use $\Omega \times S^1$ to build wrinkled fibrations over S^2 by suitably filling in the two boundary components with $\Sigma_g \times D^2$ and $\Sigma_{g-1} \times D^2$ such that the fibration structures on the boundary extends. Using the handle decomposition from [Baykur 2009] it is easy to see that this construction produces the following total spaces: $P \# S^1 \times S^3$ where P is any Σ_{g-1}-bundle over S^2 and, if $g = 1$, S^4 and some other manifolds with finite cyclic fundamental group; see [Baykur and Kamada 2010; Hayano 2011]. Having built these maps one can then apply a flip-and-slip to obtain honest simple wrinkled fibrations. In particular, S^4 carries a simple wrinkled fibration of genus two.

These examples show that simple wrinkled fibrations can be considered as a common generalization of surface bundles and (achiral) Lefschetz fibrations. The vastness of this generalization is indicated by the following remarkable theorem.

Theorem 3.8 [Williams 2010]. *Let X be a closed, oriented 4-manifold. Then any map $X \to S^2$ is homotopic to a simple wrinkled fibration of arbitrarily high genus.*

Remark 3.9. Williams' proof builds on the results of [Gay and Kirby 2007] which, in turn, depends on deep theorems in 3-dimensional contact topology. This somewhat unnatural dependence could be removed by refining the singularity theory based approach of [Baykur 2008] to produce maps which are injective on their critical locus.

Williams [2010] introduced a combinatorial description of simple wrinkled fibrations over S^2 in terms of what he calls *surface diagrams*. We will generalize his construction to the setting of general base surfaces.

Let $w : X \to B$ be a simple wrinkled fibration. The discussion in Section 2A shows that the critical locus $\mathcal{C}_w \subset X$ is a smoothly embedded circle and that w restricts to a topological embedding of $\mathcal{C}_w$ into B. Furthermore, the critical image $w(\mathcal{C}_w)$ separates B into two components. Indeed, if the complement were connected, then all regular fibers would be diffeomorphic. But according to the fold model, the topology of the fibers on the two sides of a fold arc must be different. In fact, since we require that all fibers are connected, the genus on one side has to be one higher than on the other side. We will call the two components of $B \setminus w(\mathcal{C}_w)$ the *higher*- and *lower-genus regions*.

We would like to understand more precisely how the topology of the fibers changes across the critical image. A *reference path* for w is an oriented, embedded arc $R \subset B$ that connects a point p_+ in the higher-genus region to a point p_- in the lower-genus region and intersects $w(\mathcal{C}_w)$ transversely in exactly one fold point. Then the *reference fibers* $\Sigma_\pm(R) := w^{-1}(p_\pm)$ over the *reference points* $p_\pm$ are closed, oriented surfaces.

Lemma 3.10. *A reference path $R \subset B$ induces a nonseparating simple closed curve $c(R) \subset \Sigma_+(R)$ which depends only on the isotopy class of R relative to its reference points and the cusps.*

Definition 3.11. The curve $c(R) \subset \Sigma_+(R)$ is called the *(fold) vanishing cycle* associated to R.

Proof. The fold model implies that $w^{-1}(R)$ is a cobordism from $\Sigma_+(R)$ to $\Sigma_-(R)$ on which w restricts to a Morse function with exactly one critical point of index 2. Thus $w^{-1}(R)$ is diffeomorphic to $\Sigma_+(R) \times [0, 1]$ with a (3-dimensional) 2-handle attached along a simple closed curve in $\Sigma_+(R) \times \{1\}$ which is canonically identified with a simple closed curve $c(R) \subset \Sigma_+(R)$. □

Next, let us look at what happens around the cusp. Let R_1 and R_2 be two reference paths for w with common reference points and assume that their interiors are disjoint. We call R_1 and R_2 *adjacent* if their union $R_1 \cup R_2$ bounds a disk in B that contains exactly one cusp.

Lemma 3.12. *Let R_1 and R_2 be adjacent reference paths. Then the vanishing cycles $c(R_1)$ and $c(R_2)$ in $\Sigma_+ := \Sigma_+(R_1) = \Sigma_+(R_2)$ are geometrically dual.*

Proof. As in the proof of Lemma 3.10 the preimages $w^{-1}(R_i)$, $i = 1, 2$, are both cobordisms from Σ_+ to Σ_-, each consisting of a 2-handle attachment along $c(R_i)$. By reversing the orientation of R_1 we can consider $w^{-1}(R_1)$ as a cobordism from Σ_- to Σ_+, now consisting of a 1-handle attachment. In this process the former attaching sphere of the 2-handle $c(R_1)$ becomes the belt sphere of the 1-handle.

Gluing $w^{-1}(R_1)$ and $w^{-1}(R_2)$ together along Σ_+ gives a cobordism from Σ_- to itself consisting of a 1-handle attachment followed by a 2-handle attachment. Now recall that a cusp models the death (or birth) of a canceling pair of Morse critical points. Hence, the attaching sphere of the 2-handle, which is $c(R_2)$, intersects the belt sphere of the 1-handle, which is $c(R_1)$, in a single point. □

Looking a bit ahead, our strategy will be to choose suitable collections of reference paths and to study simple wrinkled fibrations in terms of the induced collection of vanishing cycles. The only obstacle for doing so is the possibly complicated topology of the base surface. But this can easily be overcome by cutting the base into three pieces

$$B = B_+ \cup A \cup B_-$$

where A is a regular neighborhood of the critical image of w (diffeomorphic to an annulus) and $B_\pm$ are the closures of the complement of A. The subscript in $B_\pm$ indicates whether the surface is contained in the higher- or lower-genus region. Note that w restricts to surface bundles over $B_\pm$ and, although complicated, these form a rather well studied class of objects. Thus the interesting new part of w is the restriction $w^{-1}(A) \to A$ which is a simple wrinkled fibration over an annulus whose critical image is boundary parallel.

Definition 3.13. A simple wrinkled fibration $w : W \to A$ over an annulus A is called *annular* if its critical image is boundary parallel.

So in order to understand simple wrinkled fibrations over any base surface, it is enough to understand annular simple wrinkled fibrations and this is where twisted surface diagrams (see Definition 3.20 below) enter the picture. The remainder of this section is devoted to the proof of Theorem 1.1 stated in the introduction.

Remark 3.14. Gay and Kirby [2012] have published a result that contains Theorem 1.1 as a special case. Although their methods are somewhat similar to ours we feel that our approach is of independent interest.

We will split the proof of the theorem into the two obvious parts. The first part is the subject of Section 3A (see Proposition 3.25) where we show how assign twisted surface diagrams to annular simple wrinkled fibrations. The second part that shows how to build annular simple wrinkled fibrations from twisted surface diagrams is treated in Section 3C (see Proposition 3.31). In between, we will see in Section 3B that, just as Lefschetz fibrations, annular simple wrinkled fibrations are directly accessible via handlebody theory.

3A. ***Twisted surface diagrams of annular simple wrinkled fibrations.*** Consider an annular simple wrinkled fibration $w : W \to A$. We denote by $\partial_+ A$ and $\partial_- A$ the boundary components of the base annulus A contained in the higher- and lower-genus regions, respectively, and we let

$$\partial_\pm W = w^{-1}(\partial_\pm A).$$

We equip $\partial_+ A$ and $\partial_+ W$ with the opposite boundary orientation, so that W is an oriented cobordism from $\partial_+ W$ to $\partial_- W$.

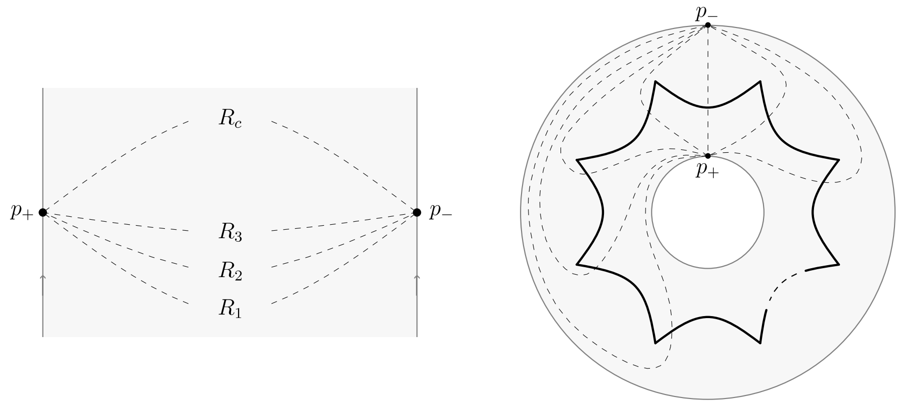

Figure 7. A reference system for an annular simple wrinkled fibration.

Definition 3.15. Let $w : W \to A$ be an annular simple wrinkled fibration. A *reference system* for w is a collection of reference paths $\mathcal{R} = \{R_1, \ldots, R_l\}$ for w (where l is the number of cusps) such that

(1) all reference paths have the same reference points $p_\pm \in \partial_\pm A$,

(2) the interiors of the arcs are pairwise disjoint,

(3) with respect to the orientations on $\partial_\pm A$ the arcs leave $\partial_+ A$ and enter $\partial_- A$ in order of increasing index (see Figure 7) and

(4) each fold arc is hit by exactly one of the R_i.

Remark 3.16. Condition (3) might need some further explanation. Assume that we have a collection of properly embedded arcs in a surface which all hit the boundary in the same point and are otherwise disjoint near that boundary component. If the boundary component is oriented, then there is a well defined notion of order for the arcs which can be described as follows. We take a small half disk around the boundary point and orient the boundary of this half disk so that it agrees with the orientation of the boundary component of the surface. For a generic choice of half disk each arc will intersect the boundary of the half disk transversely in one point and the order of these intersection points is easily seen to be independent of the choice of half disk.

As before, we denote the reference fibers by $\Sigma_\pm := \Sigma_\pm(\mathcal{R}) = w^{-1}(p_\pm)$. Using the reference fibers we can write $\partial_\pm W$ as mapping tori

$$\partial_\pm W \cong \Sigma_\pm(\mu_\pm)$$

where $\mu_\pm \in \mathrm{Mod}(\Sigma_\pm)$ is the monodromy of w over $\partial_\pm A$ (in the positive direction). We will refer to μ_+ and μ_- as the *higher* and *lower monodromies* of w.

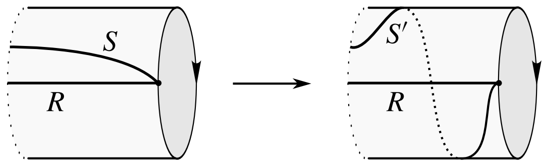

Figure 8. Swinging an arc around a boundary component.

Lemma 3.17. *Let $w : W \to A$ be an annular simple wrinkled fibration together with a reference system $\mathcal{R} = \{R_1, \dots, R_l\}$ and let $c_i = c(R_i) \subset \Sigma_+$. For $i < l$ the vanishing cycles c_i and c_{i+1} are geometrically dual, and so are $\mu_+(c_l)$ and c_1.*

In the proof we need the following construction. Let B be an oriented surface and let R and S be two properly embedded arcs in B which hit a boundary component $\partial_i B \subset \partial B$ transversely in the same point such that S enters $\partial_i B$ after R (as explained in Remark 3.16) and whose interiors are disjoint. As indicated in Figure 8 we can modify S by moving its endpoint along $\partial_i B$ resulting in a new arc S' which enters $\partial_i B$ before R and whose interior is still disjoint from R. We will say that S' is obtained from S be *swinging once around* $\partial_i B$. (Note that swinging around $\partial_i B$ is not the same as performing a boundary parallel Dehn twist since such Dehn twists are supported in the interior of B and fix a collar neighborhood of the boundary. In particular, they cannot change the order of arcs at the boundary and, moreover, in Figure 8 a boundary parallel Dehn twist applied to S would produce an arc that intersects R in its interior.)

Proof of Lemma 3.17. The first statement follows from Lemma 3.12 since for $i < l$ the reference paths R_i and R_{i+1} are clearly adjacent. For the second statement we first swing R_l once around the boundary of A so that the resulting reference path R_l' is adjacent to R_1 and thus $c(R_l')$ is geometrically dual to $c(R_1)$. Next we observe that R_l' is homotopic to R_l precomposed with the boundary curve. Thus the parallel transport along R_l' is the composition of the parallel transport along R_l and the higher-genus monodromy. In particular, we have $c(R_l') = \mu_+(c_l)$. □

Remark 3.18. Note that in the above proof we did not actually need the whole reference system but only the parts of the arcs contained in the higher-genus region.

Let us isolate the combinatorial structure encountered in the above lemma.

Definition 3.19. Let Σ be a surface. A *circuit* (of *length* l) on Σ is an ordered collection of simple closed curves $\Gamma = (c_1, \dots, c_l)$ such that any two adjacent curves c_i and c_{i+1} are geometrically dual for $i < l$. A *switch* for Γ is a mapping class $\mu \in \mathrm{Mod}(\Sigma)$ such that $\mu(c_l)$ and c_1 are geometrically dual. We say that Γ is *closed* if c_l and c_1 are geometrically dual, that is, if the identity works as a switch.

Definition 3.20. A *twisted surface diagram* is a triple $\mathfrak{S} = (\Sigma, \Gamma, \mu)$ where Σ is a closed, oriented surface, Γ is a circuit in Σ and $\mu \in \mathrm{Mod}(\Sigma)$ is a switch for Γ. In the case that Γ is a closed circuit and $\mu = \mathrm{id}$, we simply speak of *surface diagrams* and shorten the notation to $\mathfrak{S} = (\Sigma, \Gamma)$ or sometimes even $(\Sigma; c_1, \dots, c_l)$.

Remark 3.21. Note that our definition of surface diagrams is slightly different from Williams' original definition [2010]. Indeed, Williams requires that surface diagrams are induced from simple wrinkled fibrations over the sphere so that the associated annular simple wrinkled fibration has trivial higher and lower monodromies while we only require trivial higher monodromy. The reason for our deviance is that we would like to have an abstract definition of (twisted) surface diagram that does not depend on any relation to simple wrinkled fibrations. However, it turns out that the trivial lower monodromy condition for an annular simple wrinkled fibration is not easy to state in terms of its twisted surface diagrams (see Section 4A for the untwisted case) and we find it more appropriate to consider it as an extra condition.

Remark 3.22. There is no restriction on the intersections of nonadjacent curves in a circuit. Circuits in which nonadjacent curves are disjoint, so called *chains of curves*, are well known objects in the theory of mapping class groups of surfaces where they play an important role.

Remark 3.23. Sometimes it will be convenient to choose orientations on the curves in a circuit $\Gamma = (c_1, \dots, c_l)$ in order to speak of their homology classes. If the ambient surface is oriented, we always choose orientations such that the intersection of c_i and c_{i+1}, $i < l$, has positive sign.

With this terminology we can rephrase Lemma 3.17 as stating that an annular simple wrinkled fibration $w : W \to A$ together with a reference system $\mathcal{R}$ induces a twisted surface diagram

$$\mathfrak{S}_{w,\mathcal{R}} := (\Sigma_+, \Gamma_{w,\mathcal{R}}, \mu_+)$$

where the higher monodromy works as a switch.

Not surprisingly, the twisted surface diagrams constructed in Lemma 3.17 depend on the choice of the reference system. To understand this dependence we observe that a reference system is uniquely determined (up to isotopy relative to the boundary and the cusps) by specifying the first reference path — this follows directly from the definition. Furthermore, it is easy to see that any two reference paths which have the same reference points and hit the same fold arc become isotopic after suitably swinging around the boundary components of A.

Now let $\mathcal{R} = \{R_1, \dots, R_l\}$ and $\mathcal{S} = \{S_1, \dots, S_l\}$ be two reference systems with common reference points and let S_k hit the same fold arc as R_1. As in the proof of Lemma 3.17 we successively swing the arcs $S_l, S_{l-1}, \dots, S_k$ once around each boundary component to obtain a new reference system $\mathcal{S}'$ in which the first reference path hits the same fold arc as R_1. Now, by further swinging all of $\mathcal{S}'$ simultaneously, but this time independently around the boundary components, we can match the two first reference paths and thus the whole reference systems.

Let us analyze the effect of this matching procedure on the twisted surface diagram. For brevity of notation let $\mathfrak{S} = (\Sigma, \Gamma, \mu)$ be the twisted surface diagram associated to an annular simple wrinkled fibration $w : W \to A$ together with a reference system $\mathcal{R}$. Since the surface Σ and the switch μ only depend on the reference points, only the circuit $\Gamma = (c_1, \dots, c_l)$ will be affected by swinging some reference paths. Moreover, note again that the vanishing cycles c_i only depend on the part of the reference paths contained in the higher-genus region. Thus swinging around the lower-genus boundary does not change the circuit.

Now, as we have already observed, if we swing the last reference path in $\mathcal{R}$ once around both boundary components, we obtain a new reference system $\mathcal{R}'$, which induces the circuit

$$\Gamma_\mu^{[1]} := \bigl(\mu(c_l), c_1, \dots, c_{l-1}\bigr).$$

This operation of going from $\mathfrak{S}$ to $\mathfrak{S}^{[1]} := (\Sigma, \Gamma_\mu^{[1]}, \mu)$ makes sense in the abstract setting of twisted surface diagrams and we call it (and its obvious inverse) *switching*. Note that if the higher monodromy μ is trivial, then switching simply amounts to a cyclic permutation of the vanishing cycles.

Since we can relate any two reference systems for a given annular simple wrinkled fibration by suitably swinging reference paths, we see that the twisted surface diagram is well defined up to switching.

Next we want to compare the twisted surface diagrams of two equivalent annular simple wrinkled fibrations as in the commutative diagram below.

$$\begin{tikzcd} X \arrow[r, "\hat{\phi}"] \arrow[d, "w"'] & X' \arrow[d, "w'"] \\ A \arrow[r, "\check{\phi}"] & A' \end{tikzcd}$$

If $\mathcal{R}$ is a reference system for w, then $\mathcal{R}' := \check{\phi}(\mathcal{R})$ is a reference system for w'. Let $\mathfrak{S} = (\Sigma, \Gamma, \mu)$ and $\mathfrak{S}' = (\Sigma', \Gamma', \mu')$ be the associated twisted surface diagrams. Then $\hat{\phi}$ induces an orientation-preserving diffeomorphism $\phi : \Sigma \to \Sigma'$ and clearly the higher monodromies satisfy $\mu' = \phi\mu\phi^{-1}$. It is also easy to see that

$$\Gamma' = \phi(\Gamma) := \bigl(\phi(c_1), \dots, \phi(c_l)\bigr)$$

where, as usual, $\Gamma = (c_1, \dots, c_l)$. Again, the effect of an equivalence of annular simple wrinkled fibrations makes sense for abstract twisted surface diagrams and we say that $\mathfrak{S}$ and $\mathfrak{S}'$ are *diffeomorphic* via ϕ.

Combining this with switching we end up with the following definition.

Definition 3.24. Two twisted surface diagrams $\mathfrak{S}$ and $\mathfrak{S}'$ called *equivalent* if, for some integer k, $\mathfrak{S}'$ is diffeomorphic to $\mathfrak{S}^{[k]}$.

Summarizing this section so far, we have proved the first half of Theorem 1.1:

Proposition 3.25. *To an annular simple wrinkled fibration $w : W \to A$ we can assign a twisted surface diagram*

$$\mathfrak{S}_w = (\Sigma_+, \Gamma_w, \mu_+)$$

which is well defined up to equivalence. Moreover, equivalent annular simple wrinkled fibrations have equivalent twisted surface diagram.

Remark 3.26. We would like to point out that it is very convenient that only the equivalence class of the surface diagram plays a role. Indeed, in order to actually visualize the twisted surface diagram of an annular simple wrinkled fibration one has to identify the higher-genus reference fiber with some model surface and there is no canonical way to do so. However, any two such identifications will differ by a diffeomorphism of the model surface and thus be equivalent. So we can safely forget about the choice of identification whenever we are only interested in the equivalence class of the simple wrinkled fibrations or the diffeomorphism type of its total space.

3B. *Handle decompositions for annular simple wrinkled fibrations.* As a next step we relate the twisted surface diagrams associated to annular simple wrinkled fibrations to the topology of their total spaces. We will see that the situation is very similar to Lefschetz fibrations.

Proposition 3.27. *Let $w : W \to A$ be an annular simple wrinkled fibration. Then W has a relative handle decomposition on $\partial_+ W$ with one 2-handle for each fold arc. Such a handle decomposition is encoded in any twisted surface diagram for w.*

In the following we will refer to the 2-handles associated to the fold arcs as *fold handles*.

Proof. The rough idea is to parametrize A by $S^1 \times [0, 1]$ such that the composition of w and the projection $p : S^1 \times [0, 1] \to [0, 1]$ becomes a Morse function. We equip $S^1 \times [0, 1]$ with coordinates (θ, t) and refer to the direction in which t increases as *right*. We say that a parametrization $\kappa : A \to S^1 \times [0, 1]$ is *w-regular* if the critical image $C_\kappa := \kappa \circ w(\mathcal{C}_w)$ is in the following *standard position*:

- All cusps point to the right.
- Each $R_\theta := \{\theta\} \times [0, 1]$ meets C_κ in exactly one point, either in a cusp or transversely in a fold point.
- The projection p restricted to C_κ has exactly one minimum on each fold arc.

We claim that for any w-regular parametrization κ, the map

$$p_\kappa := p \circ \kappa \circ w : W \to [0, 1]$$

is a Morse function. Clearly, the critical points of p_κ are contained in $\mathcal{C}_w$. Thus we have to understand how the projection p interacts with the critical image C_κ. By the standard position assumption there are three ways how a level set $S_t := S^1 \times \{t\}$ can intersect C_κ (see Figure 9):

a) S_t intersects C_κ transversely in a fold point,
b) S_t meets C_κ in a cusp and the fold arcs surrounding the cusp are on the left side of S_t or
c) S_t is tangent to a fold arc which is located on the right side of S_t. We will refer to this phenomenon as a *left tangency*.

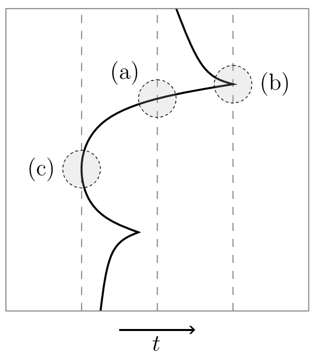

Figure 9. Level sets intersecting the critical image.

It turns out that only the left tangencies contribute critical points of p_κ. In fact, from the models for the fold and cusp we immediately see that p_κ is modeled on the compositions

$$(t, x, y, z) \mapsto (t, -x^3 + 3tx - y^2 + z^2) \mapsto t \tag{3-1}$$

in case of a cusp intersection and

$$(t, x, y, z) \mapsto (t, -x^2 - y^2 + z^2) \mapsto \pm t \tag{3-2}$$

for a transverse fold intersection (the sign depends on how the fold and cusp models are embedded) which shows that these are regular points of p_κ.

It remains to treat the concave tangencies. These occur precisely at the minima of $p_\kappa|_{C_\kappa}$. This minimum can be modeled by $t \mapsto t^2$ and it is easy to see that p_κ is modeled on

$$(t, x, y, z) \mapsto (-x^2 - y^2 + z^2 + t^2) \tag{3-3}$$

which is a Morse critical point of index 2. By assumption there is exactly one concave tangency for each fold arc and, using the correspondence between Morse functions and handle decompositions, we obtain the desired handle decomposition.

In order to understand how the fold handles are attached, consider the arcs

$$R_i := R_{\theta_i} \subset S^1 \times [0, 1],$$

where $\theta_1, \ldots, \theta_l \in S^1$ is a sequence of numbers ordered according to the orientation of S^1 (for example, the l-th roots of unity). The w-regular parametrization κ can be chosen in such a way that each R_i is a reference path for precisely one fold arc and C_κ is contained in the open annulus $S^1 \times (\epsilon, 1 - \epsilon)$ for some $\epsilon > 0$. For each R_i we obtain a vanishing cycle c_i in the fiber of w over $(\theta_i, 0) \in \partial_+ A$ and the local model for folds implies that the fold handles are attached to $\partial_+ W \times [0, \epsilon]$ along the vanishing cycles c_i pushed off into the fiber over (θ_i, ϵ) with respect to the canonical framing induced by the fiber.

The relation to twisted surface diagrams now becomes obvious. There is a canonical way to turn the reference paths $\Theta_1, \ldots, \Theta_l$ into a reference system by fixing Θ_1 and successively sliding the endpoints of the remaining arcs along the boundary onto Θ_1 against the orientation. Thus the vanishing cycles record the attaching curves of the fold handles. □

Remark 3.28. The above proposition is one of the reasons why we require the presence of cusps in the critical loci of simple wrinkled fibrations. If there were no cusps, then it would not be possible to avoid *right tangencies* which would correspond to 3-handles instead of 2-handles. Thus the presence of cusps guarantees that the total spaces of annular simple wrinkled fibrations are (relative) 2-handlebodies.

Remark 3.29. The observation that fold tangencies correspond to Morse critical points was also made by Gay and Kirby [2011a] in their more general setting of Morse 2-functions. The fact that the real part of the Lefschetz model is also a Morse function allows to include Lefschetz singularities in the discussion. Proceeding this way, one can recover Baykur's result [2009] about handle decompositions from broken Lefschetz fibrations.

Remark 3.30. The reader familiar with Lefschetz fibrations will have noticed the strong resemblance of the handle decompositions described above with the ones induced by Lefschetz fibrations. In fact, the handle decompositions have exactly the same structure except that the fold handles are attached with respect to the fiber framing while the framing of the *Lefschetz handles* differs by -1.

3C. ***Annular simple wrinkled fibrations from twisted surface diagrams.*** Using the handle decompositions exhibited in the previous section as a stepping stone we can now build annular simple wrinkled fibrations out of twisted surface diagrams and thus complete the proof of Theorem 1.1.

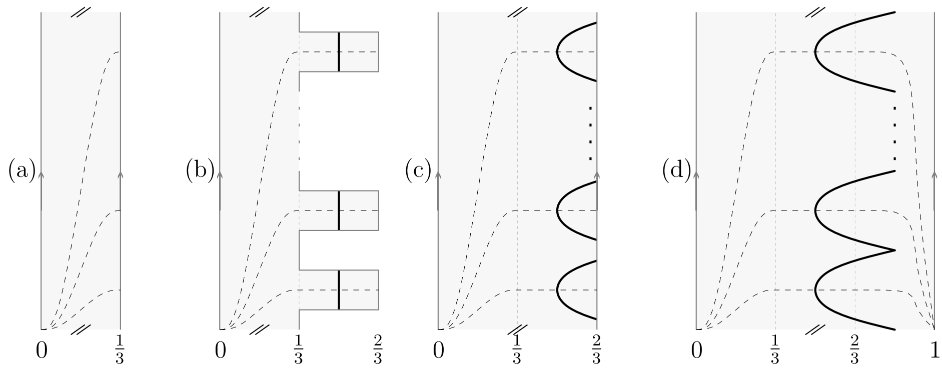

Figure 10. Building a simple wrinkled fibration from a surface diagram. Bold curves represent the critical image, and dashed curves the reference path.

Proposition 3.31. *A twisted surface diagram* $\mathfrak{S} = (\Sigma, \Gamma, \mu)$ *determines an annular simple wrinkled fibration* $w_{\mathfrak{S}} : W_{\mathfrak{S}} \to S^1 \times [0, 1]$ *with higher-genus fiber* Σ *and higher monodromy* μ.

Proof. To make the construction of $w_{\mathfrak{S}}$ more transparent we begin with some preliminary considerations. One important ingredient is the mapping cylinder $\Sigma(\mu)$ with its canonical fibration $p : \Sigma(\mu) \to S^1 = [0, 1]/\{0, 1\}$ in which we identify Σ with the fiber over $0 \sim 1$. We will also need a collection of arcs $\mathcal{R} = \{R_1, \dots, R_l\}$ in $S^1 \times [0, 1]$ that will serve as a reference system for $w_{\mathfrak{S}}$; see Figure 10(a). To construct these let $r : [0, 1] \to [0, 1]$ be a smooth function that has the constant value 1 on the interval $\left[\frac{1}{3}, \frac{2}{3}\right]$, satisfies $r(0) = r(1) = 0$ and is strictly increasing for $t \le \frac{1}{3}$ and strictly decreasing for $t \ge \frac{2}{3}$. If the length of Γ is l, then for $i = 1, \dots, l$ we let $\theta_i := (i - 1)/c$ and define

$$R_i := \left\{ \big(\theta_i r(t), t\big)/\sim \;\middle|\; t \in [0, 1] \subset S^1 \times [0, 1] \right\}.$$

We can now give the construction of $W_{\mathfrak{S}}$ and $w_{\mathfrak{S}}$ in three steps.

Step 1: We begin by taking the product $W_1 := \Sigma(\mu) \times [0, \frac{1}{3}]$ and define a map $w_1 : W_1 \to S^1 \times [0, \frac{1}{3}]$ by sending (x, t) to $(p(x), t)$.

Step 2: Next, we construct W_2 by attaching 2-handles to W_1 in the following way. Let $\Gamma = (c_1, \dots, c_l)$. Using the arc $R_i \subset S^1 \times [0, 1]$ described above we can parallel transport the curve $c_i \subset \Sigma$ to the fiber of w_1 over $\left(\theta_i, \frac{1}{3}\right)$. We attach a 2-handle to the resulting curve with respect to the fiber framing.

This choice of framing allows us to extend w_1 over each 2-handle. Indeed, we can consider attaching the i-th (4-dimensional) 2-handle as a 1-parameter family of 3-dimensional 2-handle attachments parametrized by a small neighborhood of $(\theta_i, 1)$ in $S^1 \times \{1\}$. (Of course, these neighborhoods are pairwise disjoint.) For each point θ in such a neighborhood, the restriction of w_1 to the *θ-ray* $\{\theta\} \times \left[0, \frac{1}{3}\right]$ extends to a Morse function (with one critical point of index 2) over a slightly longer ray, say $\{\theta\} \times \left[0, \frac{2}{3}\right]$, in the standard way. Using these 1-parameter families of Morse functions we can extend w_1 to map from W_2 to an annulus with "bumps" on one side, as shown in Figure 10(b), and this map has an arc of indefinite folds on each bump. We can then smooth out the bumps by standard techniques from differential topology to obtain a map $w_2 : W_2 \to S^1 \times \left[0, \frac{2}{3}\right]$ in which each 2-handle attachment has created an arc of indefinite folds whose endpoints hit the boundary of W_2 transversely in the component that was affected by the handle attachment, as in Figure 10(c); let us call this component $\partial_2 W_2$.

Step 3: For the final step we first note that the restriction of w_2 over $S^1 \times \{\frac{2}{3}\}$ is a circle valued Morse function with a pair of critical points of index 1 and 2 for each fold arc of w_2. The crucial observation is that the condition that Γ is a circuit with switch μ implies that all these pairs of critical points cancel! Thus there is a standard homotopy, which we parametrize by $\left[\frac{2}{3}, 1\right]$, from $w_2|_{\partial_2 W_2}$ to a submersion that realizes this cancellation. We let

$$W_{\mathfrak{S}} := W_2 \cup_{\partial_2 W_2} \partial_2 W_2 \times \left[\tfrac{2}{3}, 1\right]$$

and extend w_2 over the newly added collar of $\partial_2 W_2$ by tracing out the homotopy to obtain a map $w_{\mathfrak{S}} : W_{\mathfrak{S}} \to S^1 \times [0, 1]$. This last step removes all critical points from the boundary in exchange for an interior cusp for each canceling pair. Clearly $w_{\mathfrak{S}}$ is an annular simple wrinkled fibration with base diagram as in Figure 10(d).

Note that $W_{\mathfrak{S}}$ is diffeomorphic to W_2 and thus has the same relative handle decomposition. Moreover, it follows directly from the construction that $\mathcal{R}$ is a reference system for $w_{\mathfrak{S}}$ with $\mathfrak{S}$ as its twisted surface diagram. □

In order to finish the proof of Theorem 1.1 we have to show that equivalent twisted surface diagrams give equivalent annular simple wrinkled fibrations. Recall that an equivalence of surface diagrams is a combination of two things: switching and a diffeomorphism. We will treat these separately.

Lemma 3.32. *If $\mathfrak{S}$ and $\mathfrak{S}'$ are diffeomorphic, then $w_{\mathfrak{S}}$ and $w_{\mathfrak{S}'}$ are equivalent.*

Proof. Let $\mathfrak{S} = (\Sigma, \Gamma, \mu)$, $\mathfrak{S}' = (\Sigma', \Gamma', \mu')$ and let $\phi : \Sigma \to \Sigma'$ be a diffeomorphism such that $\Gamma' = \phi(\Gamma)$ and $\mu' = \phi\mu\phi^{-1}$. We will extend ϕ to a diffeomorphism $\hat{\phi} : W_{\mathfrak{S}} \to W_{\mathfrak{S}'}$ which fits in the commutative diagram

$$\begin{tikzcd} W_{\mathfrak{S}} \arrow[rr, "\hat{\phi}"] \arrow[dr, "w_{\mathfrak{S}}"'] & & W_{\mathfrak{S}'} \arrow[dl, "w_{\mathfrak{S}'}"] \\ & S^1 \times [0, 1] & \end{tikzcd}$$

This will be done by going through the steps in the proof of Proposition 3.31. Let W_i and W_i', $i = 1, 2$, denote the 4-manifolds built in each step.

From the identity $\mu' = \phi\mu\phi^{-1}$ we deduce that ϕ induces a fiber-preserving diffeomorphism $\Sigma(\mu) \to \Sigma'(\mu')$. Taking the product with the identity, we obtain $\hat{\phi}_1 : W_1 \to W_1'$.

In the second step, where the 2-handles are attached to the curves in Γ, we simply note that $\hat{\phi}_1$ maps the attaching regions into each other and can thus be extended over the 2-handles to $\hat{\phi}_2 : W_2 \to W_2'$. Note that the smoothing of the bumpy annulus does not cause any trouble since it does not involve the total space.

For the third step observe that, given a homotopy from $w_2|_{\partial_2 W_2}$ to a submersion, we can push it forward via $\hat{\phi}_2|_{\partial_2 W_2}$ to obtain such a homotopy for $w_2'|_{\partial_2 W_2'}$. □

Lemma 3.33. *If $\mathfrak{S}$ is a twisted surface diagram, then $w_{\mathfrak{S}}$ and $w_{\mathfrak{S}^{[1]}}$ are equivalent.*

Proof. If we take the canonical reference system for $w_{\mathfrak{S}}$ and swing the last reference path once around the boundary, we obtain a reference system that induces $\mathfrak{S}^{[1]}$. Thus $w_{\mathfrak{S}}$ and $w_{\mathfrak{S}^{[1]}}$ are essentially the same annular simple wrinkled fibration. □

3D. *Gluing ambiguities.* Recall that simple wrinkled fibrations over arbitrary base surfaces can be obtained from annular ones by gluing suitable surface bundles to the boundary components. To be precise, let $w_0 : W \to A$ be an annular simple wrinkled fibration and let $\pi_\pm : Y_\pm \to B_\pm$ be surface bundles over surfaces $B_\pm$ such that there are boundary components $C_\pm \subset B_\pm$ and fiber-preserving diffeomorphisms

$$\psi_\pm : \pi_\pm^{-1}(C_\pm) \to \partial_\pm W.$$

Then we can form a simple wrinkled fibration

$$w : Y_+ \cup_{\psi_+} W \cup_{\psi_-} Y_- \longrightarrow B_+ \cup_{C_+} A \cup_{C_-} B_-.$$

Of course, different choices of gluing diffeomorphisms may lead to inequivalent simple wrinkled fibrations. If we fix a pair $\psi_\pm$ of gluing maps, then we can obtain any other such pair by composing with automorphisms (in the sense of Section 2B) of the boundary fibrations $w_0 : \partial_\pm W \to S^1$. Obviously, isotopic gluing maps give rise to equivalent simple wrinkled fibrations and the gluing ambiguities are a priori parametrized by

$$\pi_0\big(\mathrm{Aut}(\partial_+ W, w)\big) \times \pi_0\big(\mathrm{Aut}(\partial_- W, w)\big).$$

However, it turns out that the first factor can be eliminated.

Lemma 3.34. *Let $w : W \to A$ be an annular simple wrinkled fibration. Then any fiber-preserving diffeomorphism of $\partial_+ W$ extends to a self-equivalence of w.*

Proof. By Theorem 1.1 we can assume that w is built from a twisted surface diagram $\mathfrak{S} = (\Sigma, \Gamma, \mu)$ such that $\partial_+ W = \Sigma(\mu)$. According to (2-6) there are two

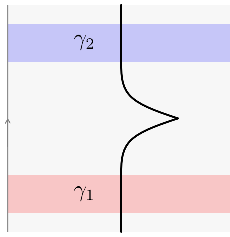

Figure 11. The relevant regions for extending nonconstant automorphisms.

types of automorphisms of $\Sigma(\mu)$, the *constant* ones coming from $C_{\mathrm{Mod}(\Sigma)}(\mu)$ and the *nonconstant* ones originating from $\pi_1(\mathrm{Diff}(\Sigma), \mathrm{id})$. The statement that constant automorphisms of $\partial_+ W$ extend to self-equivalences of w is just a reformulation of Lemma 3.32. Thus it remains to treat the nonconstant ones.

By Theorem 2.10 these only occur when Σ has genus one; we can thus assume that $\Sigma = T^2$. A refinement of Theorem 2.10 states that the map

$$\pi_1\big(\mathrm{Diff}(T^2), \mathrm{id}\big) \to \pi_1(T^2, x), \tag{3-4}$$

which sends an isotopy to the path traced out by a base point $x \in T^2$ during that isotopy, is an isomorphism; see [Earle and Eells 1969]. Note the first two curves in Γ, say c_1 and c_2, generate the fundamental group of T^2. Hence, we only have to extend the automorphisms coming from generators of $\pi_1\big(\mathrm{Diff}(T^2), \mathrm{id}\big)$ that map to c_1 and c_2 in (3-4). If we parametrize the torus by $S^1 \times S^1 \subset \mathbb{C}^2$ such that $S^1 \times \{1\}$ maps to c_1 and $\{1\} \times S^1$ maps to c_2, then such generators are given by

$$h_t^{c_1}(\xi, \eta) := (e^{2\pi i t}\xi, \eta) \quad \text{and} \quad h_t^{c_2}(\xi, \eta) := (\xi, e^{2\pi i t}\eta) \qquad (t \in [0, 1])$$

and we denote the corresponding automorphisms of $\Sigma(\mu)$ by

$$\varphi_i(x, t) := \big(h_t^{c_i}(x), t\big).$$

In order to extend φ_i to $W_{\mathfrak{S}}$ we take one step back and homotope the path h^{c_i} to be constant outside the interval where the 2-handle corresponding to c_i is attached. These intervals (times [0,1]) are highlighted in Figure 11. Outside the preimage of the regions shown in Figure 11 we can simply extend φ_i as the identity. In these region, observe that $h_t^{c_i}$ fixes c_i setwise at all times, it just rotates it more and more as t increases. It is easy to see that these rotations can be extended across the 2-handles in a way that respects the fibration structure. $\square$

Remark 3.35. The genus-one case of Example 3.7 shows that this Lemma does not hold in the absence of cusps. The above proof breaks down at the point where we need the vanishing cycles to generate the fundamental group.

4. Simple wrinkled fibrations over the disk and the sphere

We now leave the general theory behind and focus on untwisted surface diagrams, that is, pairs $\mathfrak{S} = (\Sigma, \Gamma)$ where Γ is a closed circuit in Σ, which we refer to simply as surface diagrams from now on. This will not lead to confusion since we will not encounter any twisted diagrams anymore.

By the results of the previous section, a surface diagram $\mathfrak{S}$ corresponds to an annular simple wrinkled fibration whose higher-genus boundary component has trivial monodromy. In fact, the higher-genus boundary of $w_{\mathfrak{S}} : W_{\mathfrak{S}} \to S^1 \times [0, 1]$ as constructed in Proposition 3.31 is canonically identified with the trivial fibration $\Sigma \times S^1$. We can thus fill this boundary component with $\Sigma \times D^2$ using some fiber-preserving diffeomorphism of $\Sigma \times S^1$ to obtain a simple wrinkled fibration over the disk. We denote the fibration obtain by gluing with the identity by $w_{\mathfrak{S}} : Z_{\mathfrak{S}} \to D^2$ or, by a slight abuse of notation, simply by $Z_{\mathfrak{S}}$ with the map to the disk implicitly understood. Since the boundary of the disk is contained in the lower-genus region, we refer to such fibrations as *descending* simple wrinkled fibrations (over the disk). According to Lemma 3.34, different gluing diffeomorphisms give rise to simple wrinkled fibrations equivalent to $Z_{\mathfrak{S}}$. We have thus established the following:

Proposition 4.1. *There is a bijective correspondence between the respective equivalence classes of* (*untwisted*) *surface diagrams and descending simple wrinkled fibrations over the disk.*

To make the connection to simple wrinkled fibrations over S^2, recall that by construction the boundary of $Z_{\mathfrak{S}}$ fibers over the circle. For the moment, let us say that $\mathfrak{S}$ has *trivial monodromy* if this boundary fibration is trivial (this will be made more precise in Definition 4.4 below). In this situation we can *close off* to a simple wrinkled fibration over S^2 by proceeding as above. More precisely, for a fixed boundary fiber Σ' in $Z_{\mathfrak{S}}$ we can choose a fiber-preserving diffeomorphism $\varphi : \Sigma' \times S^1 \to \partial Z_{\mathfrak{S}}$ and fill the boundary of $Z_{\mathfrak{S}}$ with a copy of $\Sigma' \times D^2$. The result is a closed 4-manifold $X^{\varphi}_{\mathfrak{S}} = Z_{\mathfrak{S}} \cup_{\varphi} \Sigma' \times D^2$ equipped with a simple wrinkled fibration over S^2 which we denote by $w^{\varphi}_{\mathfrak{S}}$. Unfortunately, this gluing process is more delicate. The main problem is that there is no canonical choice for φ; moreover, if the genus of Σ' is low, then different choices can lead to inequivalent fibrations. Combining Proposition 4.1 with the discussion in Section 3D and Theorem 2.10 leads to the cleanest possible statement:

Corollary 4.2. *Let $g > 0$ be a positive integer.*

(1) *For $g \geq 3$ there is a one-to-one correspondence between equivalence classes of genus g surface diagrams with trivial monodromy and genus g simple wrinkled fibrations over S^2.*

(2) *For $g = 2$ (respectively $g = 1$) the set of equivalence classes of genus g simple wrinkled fibrations over S^2 with equivalent surface diagrams admits a transitive action of $\mathbb{Z} \oplus \mathbb{Z}$ (respectively $\mathbb{Z}_2$).*

Recall that, according to Theorem 3.8, we can obtain all closed, oriented, smooth 4-manifolds from surface diagrams by the above process. It is thus of great interest to understand which surface diagrams have trivial monodromy and actually describe closed 4-manifolds. The following example indicates that most surface diagrams will not have trivial monodromy.

Example 4.3. Let Σ be a closed, orientable surface together with a mapping class $\phi \in \mathrm{Mod}(\Sigma)$. Then any factorization of μ into positive Dehn twists yields a Lefschetz fibration over the disk whose boundary can be identified with the mapping torus $\Sigma(\phi) = (\Sigma \times [0, 1])/(x, 1) \sim (\phi(x), 0)$. As in Example 3.6 we can turn this Lefschetz fibration into a descending simple wrinkled fibration without changing the boundary. Thus any surface bundle over the circle (with closed fibers) bounds some descending simple wrinkled fibration over the disk and any mapping class can be realized as the monodromy of a surface diagram.

In fact, the situation is very similar to the theory of Lefschetz fibrations. Any word in positive Dehn twists (or, equivalently, a finite sequence of simple closed curves) on a closed, oriented surface determines a Lefschetz fibration over the disk, the boundary fibers over the circle with monodromy being given by the product of the Dehn twists; and if this monodromy is trivial, one can close off to a Lefschetz fibration over S^2. Just as an arbitrary product of Dehn twists will not be isotopic to the identity, so will a surface diagram not give rise to a simple wrinkled fibration over S^2. The advantage of the Lefschetz setting is the direct control over the boundary.

4A. ***The monodromy of a surface diagram.*** In order to obtain a more intrinsic description of the boundary of $Z_{\mathfrak{S}}$ in terms of $\mathfrak{S}$ we need a little detour. Let $a, b \subset \Sigma$ be a pair of simple closed curves in a surface Σ that intersect transversely in a single point. We denote by Σ_a and Σ_b the surfaces obtained by surgery on the curves a and b, respectively. To be concrete, we fix tubular neighborhoods νa and νb and consider Σ_a (respectively Σ_b) as the result of filling in the two boundary components of $\Sigma \setminus \nu a$ (respectively $\Sigma \setminus \nu b$) with disks. We can assume that $\nu(a \cup b) := \nu a \cup \nu b$ is diffeomorphic to a once punctured torus — for convenience we also assume that it has a smooth boundary in Σ. Observe that $\Sigma \setminus \nu(a \cup b)$ has one boundary component and is contained in both Σ_a and Σ_b as a subsurface. Furthermore, the closure of $\nu b \setminus \nu a$ (respectively $\nu a \setminus \nu b$) is a disk in Σ_a (respectively Σ_b). It follows that, up to isotopy, there is a unique diffeomorphism

$$\kappa_{a,b} : \Sigma_a \to \Sigma_b$$

which restricts to the identity on $\Sigma \setminus \nu(a \cup b)$. Furthermore, we can assume that $\kappa_{a,b}$ maps $b \setminus \nu a$ onto $a \setminus \nu b$.

Now let $\mathfrak{S} = (\Sigma; c_1, \ldots, c_l)$ be a surface diagram and consider the associated simple wrinkled fibration $w_{\mathfrak{S}} : Z_{\mathfrak{S}} \to D^2$. Then each adjacent pair of curves c_i and c_{i+1} fits the above situation and we thus get a collection of diffeomorphisms

$$\kappa_{c_i,c_{i+1}} : \Sigma_{c_i} \to \Sigma_{c_{i+1}}.$$

Moreover, it follows from the definition of surface diagrams that the composition

$$\mu_{\mathfrak{S}} := \kappa_{c_l,c_1} \circ \kappa_{c_{l-1},c_l} \circ \cdots \circ \kappa_{c_1,c_2}$$

maps Σ_{c_1} to itself and it is easy to see that its isotopy class does not depend on any of the implicit choices involved in its definition.

Definition 4.4. The mapping class $\mu_{\mathfrak{S}} \in \mathrm{Mod}(\Sigma_{c_1})$ represented by the composition above is called the *monodromy* of $\mathfrak{S}$.

This name is justified by the following lemma.

Lemma 4.5. *Let $\mathfrak{S} = (\Sigma, \Gamma)$ be a surface diagram. Then the boundary fibration $(\partial Z_{\mathfrak{S}}, w_{\mathfrak{S}})$ can be identified with the mapping torus $\Sigma_{c_1}(\mu_{\mathfrak{S}})$.*

Proof. By the construction of $w_{\mathfrak{S}}$ its fiber over the origin is naturally identified with Σ. Furthermore, recall that the annular fibration associated to $\mathfrak{S}$ is equipped with a reference system whose reference paths we can naturally extend from the annulus to the disk by connecting them to the origin. The result is a collection of reference paths $R_1, \ldots, R_l$ from the origin to the boundary of the disk and we denote its endpoints by $\theta_1 \ldots, \theta_l \in S^1$. Observe that such a reference path, say R_i, gives rise to an identification of the fiber over θ_i with the surface Σ_{c_i} obtained from surgery on c_i where c_i is the vanishing cycle associated to R_i.

Now consider the region in the base bounded by two adjacent reference paths R_i and R_{i+1}. Using a suitable notion of parallel transport we see that the preimage of this region contains a trivial bundle with fiber $\Sigma \setminus \nu(c_i \cup c_{i+1})$. In particular, the parallel transport along the boundary segment from θ_i to θ_{i+1} restricts to the identity on the complement of $\nu(c_i \cup c_{i+1})$ and thus must be isotopic to $\kappa_{c_i,c_{i+1}}$ and the claim follows. □

It is also possible to describe the monodromy in terms of the original surface Σ. This takes us on another small detour. Let $a \subset \Sigma$ be a nonseparating simple closed curve in a surface Σ and let $\mathrm{Mod}(\Sigma, a)$ denote the subgroup of $\mathrm{Mod}(\Sigma)$ consisting of all elements that fix a up to isotopy. Recall that there is a short exact sequence

$$1 \longrightarrow \langle \tau_a \rangle \longrightarrow \mathrm{Mod}(\Sigma, a) \xrightarrow{\mathrm{cut}_a} \mathrm{Mod}(\Sigma \setminus a) \longrightarrow 1 \tag{4-1}$$

where we consider $\Sigma \setminus a$ as a twice punctured surface (see [Farb and Margalit 2011, Chapter 3] and also [Ivanov 1992, Section 7.5] for a proof that cut_a is well defined). The complement $\Sigma \setminus a$ can be related to the surgered surface Σ_a as follows. In Σ_a there is an obvious pair of points, namely the centers of the surgery disks. If we denote by Σ_a^* the surface obtained by marking these points, then $\Sigma \setminus a$ is canonically identified (at least up to isotopy) with Σ_a^* and thus $\mathrm{Mod}(\Sigma \setminus a)$ is canonically isomorphic to $\mathrm{Mod}(\Sigma_a^*)$. Hence, we can define the *surgery homomorphism*

$$\Phi_a : \mathrm{Mod}(\Sigma, a) \to \mathrm{Mod}(\Sigma_a)$$

as the composition

$$\begin{tikzcd}
\mathrm{Mod}(\Sigma, a) \arrow[r, "\mathrm{cut}_a"'] \arrow[rrr, bend left, "\Phi_a"] & \mathrm{Mod}(\Sigma \setminus a) \arrow[r, "\cong"'] & \mathrm{Mod}(\Sigma_a^*) \arrow[r, "\mathrm{forget}"'] & \mathrm{Mod}(\Sigma_a)
\end{tikzcd}$$

where the last map is induced by forgetting the marked points in Σ_a^*.

Applying this to surface diagram we obtain the following.

Lemma 4.6. *Let $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ be a surface diagram. Then*

$$\tilde{\mu}_{\mathfrak{S}} := \tau_{\tau_{c_l}(c_1)} \circ \tau_{\tau_{c_{l-1}}(c_l)} \circ \tau_{\tau_{c_1}(c_2)} \in \mathrm{Mod}(\Sigma)$$

is contained in $\mathrm{Mod}(\Sigma, c_1)$ *and satisfies* $\Phi_{c_1}(\tilde{\mu}_{\mathfrak{S}}) = \mu_{\mathfrak{S}}$.

Proof. Since c_i and c_{i+1} are geometrically dual, the mapping class $\tau_{\tau_{c_{l-1}}(c_l)}$ has a representative $T \in \mathrm{Diff}^+(\Sigma)$ that maps c_i to c_{i+1} (as a set). The claim then follows from the observation that the diagram

$$\begin{tikzcd}
\Sigma \arrow[d, "\tau_{\tau_{c_i}(c_{i+1})}"'] & \Sigma \setminus c_i \arrow[l] \arrow[r] \arrow[d, "T"] & \Sigma_{c_i}^* \arrow[d, "\kappa_{c_i, c_{i+1}}"] \\
\Sigma & \Sigma \setminus c_{i+1} \arrow[l] \arrow[r] & \Sigma_{c_{i+1}}^*
\end{tikzcd}$$

commutes up to isotopy. □

The above makes it interesting to study the map Φ_{c_1} and its kernel.

Lemma 4.7. *Let $a \subset \Sigma$ be a nonseparating simple closed curve. The group* $\mathrm{Mod}(\Sigma, a)$ *is generated by elements of the form τ_c and $\Delta_{a,b} := (\tau_a \tau_b)^3$, where $i(a, c) = 0$ and $i(a, b) = 1$.*

We refer to the mapping classes $\Delta_{a,b}$ as Δ-*twists*. Note that Δ-twists are defined for arbitrary pairs of geometrically dual curves and do not have to involve the curve a in the above Lemma.

Proof. It follows from the short exact sequence (4-1) that we can obtain a generating set for $\mathrm{Mod}(\Sigma, a)$ by lifting a generating set for $\mathrm{Mod}(\Sigma \setminus a)$ and adding the Dehn twist about a. As a generating set for $\mathrm{Mod}(\Sigma \setminus a)$ we can take the collection Dehn twists and so called *half-twists* about simple arcs connecting the two punctures. Then the Dehn twists in $\mathrm{Mod}(\Sigma \setminus a)$ have obvious lifts in $\mathrm{Mod}(\Sigma)$ and it is easy to see that each half-twist lifts to a Δ-twist. $\square$

Corollary 4.8. *The kernel of* $\Phi_a : \mathrm{Mod}(S, a) \to \mathrm{Mod}(\Sigma_a)$ *contains the Dehn twist* τ_a *as well as all* Δ*-twists involving* a.

The expert will have noticed that the mapping class $\tilde{\mu}_{\mathfrak{S}}$ in Lemma 4.6 is simply the monodromy of the boundary of the Lefschetz part of the simplified broken Lefschetz fibration obtained from $w_{\mathfrak{S}}$ by unsinking all the cusps. Of course, there are many different lifts of $\mu_{\mathfrak{S}}$ to $\mathrm{Mod}(\Sigma)$. For example, it follows from the braid relations for the pairs of adjacent curves that

$$\begin{aligned}\tilde{\mu}_{\mathfrak{S}} &= \tau_{c_1}^{-c}(\tau_{c_l}\tau_{c_1})(\tau_{c_{l-1}}\tau_{c_l})\dots(\tau_{c_1}\tau_{c_2})\\ &= \tau_{c_1}^{-2c}(\tau_{c_l}\tau_{c_1}\tau_{c_l})(\tau_{c_{l-1}}\tau_{c_l}\tau_{c_{l-1}})\dots(\tau_{c_1}\tau_{c_2}\tau_{c_1})\end{aligned}$$

and since τ_{c_1} is contained in the kernel of Φ_{c_1} we obtain two other choices. To illustrate these mapping class group techniques we produce some examples of surface diagrams with trivial monodromy.

Example 4.9. Given a not necessarily closed circuit $\Gamma = (c_1, \dots, c_l)$ in an oriented surface Σ we can form a closed circuit $D\Gamma := (c_1, \dots, c_{l-1}, c_l, c_{l-1}, \dots, c_2)$ which we call the *double of* Γ. We claim that the surface diagram $D\mathfrak{S} := (\Sigma, D\Gamma)$ has trivial monodromy. For convenience let us write $\tau_i = \tau_{c_i}$. As explained above the monodromy of $D\mathfrak{S}$ can be lifted to $\mathrm{Mod}(\Sigma)$ as

$$\begin{aligned}\mu &= (\tau_2\tau_1\tau_2)\dots(\tau_{l-2}\tau_{l-1}\tau_{l-2})(\tau_{l-1}\tau_l\tau_{l-1})(\tau_l\tau_{l-1}\tau_l)(\tau_{l-1}\tau_{l-2}\tau_{l-1})\dots(\tau_1\tau_2\tau_1)\\ &= (\tau_2\tau_1\tau_2)\dots(\tau_{l-2}\tau_{l-1}\tau_{l-2})\Delta_{c_{l-1},c_l}(\tau_{l-1}\tau_{l-2}\tau_{l-1})\dots(\tau_1\tau_2\tau_1).\end{aligned}$$

Our goal is to factor this expression into a sequence of Δ-twists involving c_1. The key observation is that

$$\begin{aligned}&(\tau_{l-2}\tau_{l-1}\tau_{l-2})\ \Delta_{c_{l-1},c_l}\ (\tau_{l-1}\tau_{l-2}\tau_{l-1})\\ &= (\tau_{l-2}\tau_{l-1}\tau_{l-2})\ \Delta_{c_{l-1},c_l}\ (\tau_{l-2}\tau_{l-1}\tau_{l-2})\\ &= (\tau_{l-2}\tau_{l-1}\tau_{l-2})\ \Delta_{c_{l-1},c_l}\ (\tau_{l-2}\tau_{l-1}\tau_{l-2})^{-1}\ \Delta_{c_{l-2},c_{l-1}}\\ &= \Delta_{\tau_{l-2}\tau_{l-1}\tau_{l-2}(c_{l-1}),\tau_{l-2}\tau_{l-1}\tau_{l-2}(c_l)}\ \Delta_{c_{l-2},c_{l-1}}\\ &= \Delta_{c_{l-2},\tau_{l-2}\tau_{l-1}\tau_{l-2}(c_l)}\ \Delta_{c_{l-2},c_{l-1}}.\end{aligned}$$

Applying this repeatedly we eventually obtain

$$\mu = \Delta_{c_1,\delta_l}\Delta_{c_1,\delta_{l-1}}\dots\Delta_{c_1,\delta_2},$$

where $\delta_k := \tau_1\tau_2\tau_1 \dots \tau_{k-2}\tau_{k-1}\tau_{k-2}(c_k)$. Hence, the monodromy of $D\mathfrak{S}$ is trivial by Corollary 4.8.

It is also possible to show that $D\mathfrak{S}$ has trivial monodromy by directly constructing a simple wrinkled fibration over S^2. This construction will also justify our terminology. The key observation is that, even if Γ is not closed, the ideas in the proof of Proposition 3.31 can be used to build a wrinkled fibration over the disk.

Indeed, by attaching 2-handles to $\Sigma \times D^2$ along the fiber framed curves c_i in boundary fibers ordered according to the orientation of S^1 we obtain a 4-manifold with boundary P_Γ together with a map to the disk which has an arc of folds for each 2-handle and each arc gives rise to a pair of Morse critical points on the boundary. As in the third step of the proof of Proposition 3.31 we can trade pairs of critical points on the boundary coming from c_i and c_{i+1}, $i < l$, for cusps in the interior. What remains is a wrinkled fibration on P_Γ over the disk with two critical points on the boundary, one coming from c_1 and the other from c_l. Of course, if Γ is closed, then P_Γ is diffeomorphic to $Z_\mathfrak{S}$ where $\mathfrak{S} = (\Sigma, \Gamma)$, but the corresponding map to the disk is different.

If we apply this construction to the reversed circuit $\overline{\Gamma} = (c_l, \dots, c_1)$, then we obtain another 4-manifold $P_{\overline{\Gamma}}$ and it is easy to see that the self-diffeomorphism of $\Sigma \times D^2$ which sends (p, x) to $(p, -x)$ induces an orientation-preserving diffeomorphism from $P_{\overline{\Gamma}}$ to $\overline{P}_\Gamma$. We thus obtain a wrinkled fibration on $\overline{P}_\Gamma$ and the identity map of ∂P_Γ provides an orientation-reversing and fiber-preserving diffeomorphism of the boundary fibrations on P_Γ and $\overline{P}_\Gamma$. Hence, the fibrations on P_Γ and $\overline{P}_\Gamma$ give rise to a wrinkled fibration over S^2 on the double $DP_\Gamma = P_\Gamma \cup_{\mathrm{id}} \overline{P}_\Gamma$ which turns out to be a simple wrinkled fibration with surface diagram $D\mathfrak{S}$.

4B. ***Drawing Kirby diagrams.*** In this section we show how to translate surface diagrams into Kirby diagrams of the associated simple wrinkled fibrations. For the necessary background we refer the reader to [Gompf and Stipsicz 1999]. Throughout, we use Akbulut's *dotted circle notation* for 1-handles to avoid ambiguities for framing coefficients.

Descending simple wrinkled fibrations. Let $w_\mathfrak{S} : Z_\mathfrak{S} \to D^2$ be a descending simple wrinkled fibration of genus g with surface diagram $\mathfrak{S} = (\Sigma_g; c_1, \dots, c_l)$. Recall that the associated handle decomposition of Z is obtained from (some handle decomposition of) $\Sigma_g \times D^2$ by attaching 2-handles along $c_i \subset \Sigma_g \times \{\theta_i\}$ with the fiber framing where $\theta_1, \dots, \theta_l \in S^1$ are ordered according to the orientation on S^1. So in order to draw a Kirby diagram for $Z_\mathfrak{S}$ we need to find a diagram for $\Sigma \times D^2$ in which the fibers of the boundary should be as clearly visible as possible.

A convenient choice is the diagram shown in Figure 12 which is induced from the obvious handle decomposition of Σ_g with one 0-handle, $2g$ 1-handles and one 2-handle. One fiber of $\Sigma_g \times S^1$, which we identify with Σ_g, is clearly visible and

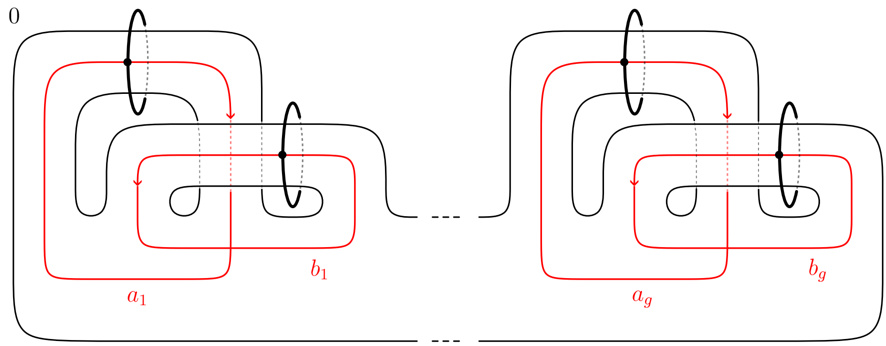

Figure 12. A diagram for $\Sigma_g \times D^2$ where fiber and blackboard framing agree. The red curves show a basis for $H_1(\Sigma_g)$.

the canonical generators $a_1, b_1, \dots, a_g, b_g$ for $H_1(\Sigma_g)$ are also indicated. We have chosen the orientations such that $\langle a_i, b_i\rangle_{\Sigma_g} = 1$. Another advantage of this picture is that the fiber framing agrees with the blackboard framing. One minor drawback is that the picture does not immediately show *all* fibers of $\Sigma_g \times S^1$ but only an interval worth of them (just thicken the surface a little). However, this is actually enough for our purposes since we only need the fibers over the interval $[\theta_1, \theta_l] \subset S^1$. To get the orientations right we require that the orientation of the fiber agrees with the standard orientation of the plane and, according to the "fiber first convention", the positive S^1-direction points toward the reader.

With this understood, it is easy to locate the attaching curves of the fold handles in the diagram and it remains to determine their framing coefficients. More generally, we can describe the linking form of the link corresponding to the fold handles. It should be no surprise that the framing and linking information in the diagram depends on our choice of the handle decomposition for Σ_g.

Let $c \subset \Sigma_g$ be a simple closed curve. After choosing an orientation its homology class $[c] \in H_1(\Sigma)$ can be expressed as

$$[c] = \sum_{i=1}^{g} \big(n_{a_i}(c)\, a_i + n_{b_i}(c)\, b_i\big).$$

We identify Σ_g with $\Sigma_g \times \{0\}$ and, by a slight abuse of notation, we continue to denote the canonical push-off of c to $\Sigma_g \times \{z\}$, $z \in D^2$, by c.

Lemma 4.10. *For a simple closed curve $c \subset \Sigma_g \times \{\theta\}$, $\theta \in [\theta_1, \theta_l] \subset S^1$, the framing coefficient of the fiber framing in Figure 12 is given by*

$$\mathrm{fr}(c) = \sum_{i=1}^{g} n_{a_i}(c) n_{b_i}(c). \tag{4-2}$$

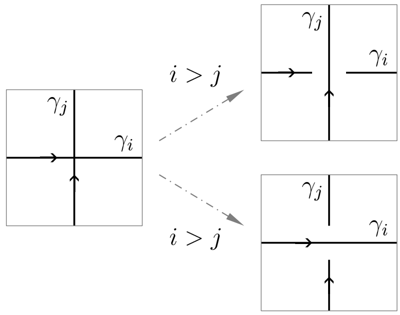

Figure 13. An intersection in a surface diagram and its crossing in the Kirby diagram.

Furthermore, if $c \subset \Sigma_g \times \{\theta\}$ and $c' \subset \Sigma_g \times \{\theta'\}$ are two oriented simple closed curves, with $\theta, \theta' \in [\theta_1, \theta_l]$, their linking number in Figure 12 is

$$\operatorname{lk}(c, c') = \tfrac{1}{2}\operatorname{sgn}(\theta - \theta')\langle c, c'\rangle + \tfrac{1}{2}\sum_{i=1}^{g}\bigl(n_{a_i}(c)n_{b_i}(c') + n_{a_i}(c')n_{b_i}(c)\bigr), \tag{4-3}$$

where $\langle c, c'\rangle$ is the algebraic intersection number of c and c' in Σ_g and sgn *denotes the sign of a real number.*

Proof. First observe that $c \subset \Sigma_g \times \{\theta\}$ can be isotoped off the 2-handle of Σ_g so that it becomes completely visible in Figure 12 and, since the fiber framing and blackboard framing agree, its framing coefficient is given by its writhe in the diagram — the signed count of crossings with some chosen orientation. From the way the diagram is drawn it is clear that each crossing is caused by c running over a_i *and* b_i for some i and that their signed sum is given by the right side of (4-2).

The statement about linking numbers follows from a similar count of crossings. Recall that the linking number of two oriented knots can be computed from any link diagram as half of the signed number of crossings. The second term on the right side of (4-3) arises just as above. However, the first term deserves some explanation. Each (transverse) intersection point of c and c' in Σ_g contributes a crossing in the diagram. Now, the sign of the crossing depends on two things: the sign of the intersection point and the information which strand is on top in the diagram. From Figure 13 we see that the contribution of each crossing is exactly as in (4-2). □

Remark 4.11. Formula (4-3) can be used to obtain a description of the intersection form of the 4-manifold $Z_{\mathfrak{S}}$ using only the data in $\mathfrak{S}$. Also, since (4-3) only depends on the homology classes of the curves in $\mathfrak{S}$, so do the intersection form and, in particular, the signature of $Z_{\mathfrak{S}}$. We will return to this observation in a future publication.

The diagrams of simple wrinkled fibrations derived from Figure 12 are good for abstract reasoning, however, in practice it is convenient to start with a cleaner

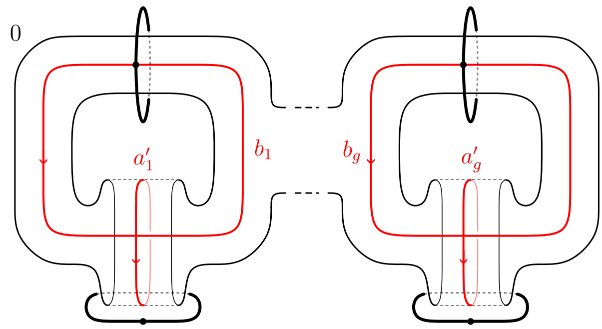

Figure 14. A cleaner diagram of $\Sigma_g \times D^2$.

diagram for $\Sigma_g \times D^2$ such as the one shown in Figure 14. In this picture, the fiber appears as the boundary sum of regular neighborhoods of the basis curves $\{a'_i, b_i\}_{i=1}^g$ which, in turn, appear as meridians to the dotted circles. The framing coefficient of the fiber framing for simple closed curves on a fiber in Figure 14 can be computed as follows. It is not hard to see that Figure 14 is obtained from Figure 12 by a sequence of 1-handle slides and an isotopy of the 2-handle and vice versa. Note that these moves do not change the framing coefficients of any other 2-handles that might have been around. Moreover, during the moves, the b-curves remain fixed, while the a-curves undergo some changes. When pulling a'_i in Figure 14 back to Figure 12 one obtains a representative for the element

$$[a_1, b_1] * \cdots * [a_{i-1}, b_{i-1}] * a_i \in \pi_1(\Sigma_g)$$

where $[x, y] = xyx^{-1}y^{-1}$. The important observation is that while this curve is not isotopic to a_i it does represent the same homology class. As a consequence, formula (4-2) can be used for Figure 14 with a_i replaced by a'_i.

Closing off and the last 2-handle. Recall that our motivation comes from Williams' theorem that all closed, oriented 4-manifolds admit simple wrinkled fibrations over S^2. We have seen that these can be described (up to equivalence) by surface diagrams with trivial monodromy and we have already mentioned that it is in general not easy to check whether the monodromy of a given surface diagram is trivial. But the situation is even worse. Say that we know for some reason that a given surface diagram has trivial monodromy and let us also assume that the genus is at least three so that there are no gluing ambiguities. Even in this case it is not clear at all how the surface diagram encodes the information to complete the Kirby diagram.

To be more precise, let $w : X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S}$. Let $\nu\Sigma_-$ be a neighborhood of a lower-genus fiber and let $Z := X \setminus \nu\Sigma_-$. Then w restricts to a descending simple wrinkled fibration on Z and ∂Z can be identified with $\Sigma_- \times S^1$ so that $\mathfrak{S}$ must have trivial monodromy. We can draw a

Kirby diagram for Z as described in the previous section and to complete it to a diagram for X we have to understand how to glue $\nu\Sigma_-$ back in.

We can choose a handle decomposition for $\nu\Sigma_-$ with one 0-handle, $2g(\Sigma_-)$ 1-handles and one 2-handle. Turning this upside down results in a relative handle decomposition on $\partial Z \cong \Sigma_- \times S^1$ with one 2-handle, $2g(\Sigma_-)$ 3-handles and a 4-handle. The general theory tells us that the 3- and 4-handles attach in a standard way once we know how to attach the 2-handle. Unfortunately, it turns out to be rather difficult to locate this *last 2-handle* in the Kirby diagram for Z.

Our knowledge about the last 2-handle is a priori limited to the following observation. If we identify $\nu\Sigma_-$ with $\Sigma_- \times D^2$, then the attaching curve of the last 2-handle corresponds to $\{p\} \times \partial D^2$ for some $p \in \Sigma_-$. In particular, we see that it must be attached along a section of the boundary fibration $(\partial Z, w)$.

Remark 4.12. Given a surface diagram $\mathfrak{S}$ with trivial monodromy there is a general method for finding possible last 2-handles for $Z_{\mathfrak{S}}$ which is not very conceptual but still useful in some situations. One considers a Kirby diagram for $Z_{\mathfrak{S}}$ as a surgery diagram for $\partial Z_{\mathfrak{S}}$ and performs (3-dimensional) Kirby moves until the fibration structure is clearly visible as $\Sigma_- \times S^1$. In such a diagram it is easy to locate attaching curves for possible last 2-handles which one can then pull back to the original diagram by undoing the moves and dragging the curves along. This strategy also works for Lefschetz fibrations as discussed in [Gompf and Stipsicz 1999, Chapter 8.2].

Just as in the Lefschetz case, the situation becomes easier if one knows that $Z_{\mathfrak{S}}$ can be closed off to a fibration over S^2 which admits a section. The proof of the following lemma is the same as in the Lefschetz case and we refer the reader to [Gompf and Stipsicz 1999].

Lemma 4.13. *Let $w : X \to S^2$ be a simple wrinkled fibration with surface diagram $\mathfrak{S}$. If w admits a section of self-intersection k, then the last two handle appears in the diagram for $Z_{\mathfrak{S}}$ as a k-framed meridian of the* 2*-handle corresponding to the fiber. Furthermore, if $\mathfrak{S}$ is a surface diagram and a meridian as above can be used to attach the last* 2*-handle, then the corresponding simple wrinkled fibration admits a section of self-intersection k.*

In order to illustrate Remark 4.12 and Lemma 4.13 as well as our method of drawing Kirby diagrams we give an example which is also a warm-up for the next section.

Example 4.14. Let $a, b \subset \Sigma_g$ be a geometrically dual pair of simple closed curves. We claim that $\mathfrak{S} = (\Sigma_g; a, \tau_b(a), b)$ is a surface diagram for $\Sigma_{g-1} \times S^2 \# \overline{\mathbb{C}P^2}$. We can assume that a and b are the standard generators a_1 and b_1 in Figure 14 and Figure 15 shows the final Kirby diagram. In order to see how we got there let us first

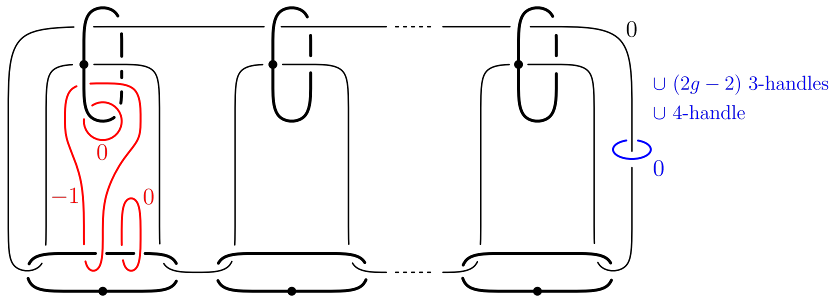

Figure 15. Manifolds with surface diagram $(\Sigma_g; a, \tau_b(a), b)$.

ignore all the blue components. What is left is just the Kirby diagram for $Z_{\mathfrak{S}}$. The framings on the fold handles can either be computed using Lemma 4.10 (together with Proposition 2.7) or by hand (the curve is simple enough to draw a parallel push-off in the fiber direction and compute the linking number). We now perform the obvious handle moves: using the meridians to the two 1-handles on the left we first unlink the -1-framed fold handle (corresponding to $\tau_b(a)$) to obtain a -1-framed unknot isolated from the rest of the diagram, then we unlink the black 2-handle (corresponding to the fiber) and finally cancel the 1-handles and their meridians. Obviously, the thus obtained diagram shows $\Sigma_{g-1} \times D^2 \# \overline{\mathbb{C}P^2}$ and the boundary is clearly visible as $\Sigma_{g-1} \times S^1$. Moreover, it is easy to see that the last 2-handle can be attached along a 0-framed meridian to the fiber 2-handle and the resulting manifold is $\Sigma_{g-1} \times S^2 \# \overline{\mathbb{C}P^2}$ as claimed. Finally, since we attached the last 2-handle in a region that was not affected by the Kirby moves it will not change when we undo the moves again and we arrive at Figure 15. Lemma 4.13 then tells us that the corresponding simple wrinkled fibration will have a section of self-intersection zero.

Note that for $g \geq 3$ any other choice for the last 2–handle that might have been possible leads to an equivalent fibration whose total space is diffeomorphic to $\Sigma_{g-1} \times S^2 \# \overline{\mathbb{C}P^2}$. In the lower-genus cases there are more options. However, in any case one will end up with a blow-up of some surface bundle over S^2.

4C. ***Relation to broken Lefschetz fibrations.*** Let $w : X \to B$ be a simple wrinkled fibration. After trading all the cusps for Lefschetz singularities by applying Lekili's unsinking modification we obtain a broken Lefschetz fibration

$$\beta_w : X \to B$$

with one round singularity, smoothly embedded in the base, and all its Lefschetz points on the higher-genus side. If the base is the sphere or the disk, then β_w is a *simplified broken Lefschetz fibration* in the sense of [Baykur 2009] and thus induces another handle decomposition of X.

In order to relate these two handle decompositions, let us briefly review how a simplified broken Lefschetz fibration $\beta : X \to B$ gives rise to a handle decompositions. Much in the spirit of simple wrinkled fibrations one chooses a reference point in the higher-genus region together with a collection of disjointly embedded arcs $L_1, \dots, L_k, R \subset B$, where k is the number of Lefschetz singularities, emanating from the reference point such that each L_i ends in a Lefschetz point and R passes through the round singularity once. Such a system of arcs is known as a *Hurwitz system* for β. The arcs in a Hurwitz system then give rise to simple closed curves in the reference fiber Σ to which we shall refer to as the *Lefschetz vanishing cycles* $\lambda_1, \dots, \lambda_k \subset \Sigma$ and the *round vanishing cycle* ρ. A handle decomposition of X is then given as follows:

- Start with $\Sigma \times D^2$.
- Going around S^1, attach a *Lefschetz handle* along the λ_i pushed off into fibers over S^1 (that is, 2-handles with framing -1 with respect to the fiber framing).
- Attach a *round* 2*-handle* along ρ.

The round 2-handle decomposes into a 2-handle and a 3-handle such that the 3-handle goes over the 2-handle geometrically twice and the 2-handle is attached along ρ with respect to the fiber framing. (For more details see [Baykur 2009].)

Now let $w : X \to B$ be a simple wrinkled fibration and let β_w be the associated simplified broken Lefschetz fibration. Given a reference system $\mathcal{R} = \{R_i\}$ for w with associated surface diagram (Σ, Γ) there is a canonical Hurwitz system for β_w. Since the unsinking homotopy is supported near the cusps we can assume that the nothing happens around the reference paths. Now observe that the arcs R_i cut the higher-genus region into triangles each containing a single Lefschetz singularity of β_w. Thus, up to isotopy, there is a unique arc L_i in the triangle bounded by R_i and R_{i+1} going from the reference fiber to the Lefschetz singularity and for the round singularity we take the arc $R = R_1$. According to Lekili [2009], the vanishing cycles of β_w with respect to this Hurwitz system are given by

$$\lambda_i = \tau_{c_i}(c_{i+1}) \quad \text{and} \quad \rho = c_1.$$

We can go from the handle decomposition induced by β_w to the one induced by w using the following handlebody interpretation of the (un)sinking deformation.

Assume that we have a Lefschetz singularity next to a fold arc that is *sinkable*, that is, the Lefschetz and fold vanishing cycles intersect in one point. (In other words, it is the resulting of unsinking a cusp.) In terms of handle decompositions the situation before and after the sinking process is locally described in Figure 16. (These handle decompositions have already appeared in a disguised form in [Lekili 2009].) Clearly, both pictures describe a 4-ball and they are related by an obvious

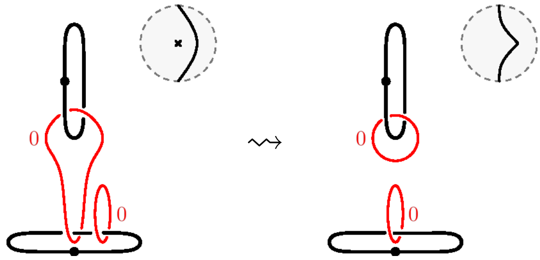

Figure 16. A Lefschetz singularity before and after sinking. (The Lefschetz 2-handle on the left runs over both 1-handles. One readily checks that it is correctly framed.)

2-handle slide. Indeed, to go from (a) to (b) one has to slide the Lefschetz handle over the fold handle in such a way that it unlinks from the lower 1-handle. Note that his handle slide is compatible with the fibration structures in the sense that the attaching curves stay on the fibers. Moreover, it mysteriously adjusts the framings exactly as needed.

Remark 4.15. Although the handle slide described above seems to be a correct interpretation of Lekili's (un)sinking deformation it is a priori not obvious why this should be true. In fact, the deformation is a combination of wrinkling, merging and flipping (see [Lekili 2009, Figure 8]) and does not seem very atomic. On the other hand, the handle slide is an atomic modification of the handlebodies. It would be interesting to see a 1-parameter family of Morse functions associated with the (un)sinking deformation that would exhibit the handle slide.

This shows that, if we start we the handle decomposition of β_w, then sliding λ_1 over $\rho = c_1$ produces a fiber framed attaching curve λ_1' which is isotopic to c_2. Successively sliding λ_i over $\lambda_{i-1}' \sim c_i$ results in fiber framed attaching curves λ_i' isotopic to c_{i+1}. Altogether we end up with fiber framed curves $\lambda_1', \dots, \lambda_l', \rho$. The final observation is that λ_l' is isotopic to $\rho = c_1$ and can be unlinked and isolated from the rest of the diagram to form a zero framed unknot which cancels the 3-handle coming from the round singularity. What we are left with is the decomposition associated to w.

5. Substitutions

Let $\mathfrak{S} = (\Sigma, \Gamma)$ be a surface diagram and let Λ be a subcircuit of Γ. If Λ' is any circuit that starts and ends with the same curves as Λ, then we can build a new surface diagram (Σ, Γ') where Γ' is obtained by replacing Λ with Λ'. We call this operation a *substitution of type* $(\Lambda|\Lambda')$. Similar substitution techniques for Lefschetz fibrations are studied in [Endo and Gurtas 2010; Endo et al. 2011].

Passing to the associated simple wrinkled fibrations one can ask how such a substitution affects the total spaces. In the following we treat two instances in which this question can be answered. Our main tools are the handle decompositions exhibited in the previous section.

Let Z be a compact 4-manifold, possibly with nonempty boundary. Recall that the operations of taking connected sums with $\overline{\mathbb{C}P^2}$ and $\Sigma^2 \times S^2$ (taken in the interior of Z) are commonly known as *blow-up* and *sum stabilization*. We will be slightly more general and also call connected sums with $\mathbb{C}P^2$ blow-ups and connected sums with $\mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$, the twisted S^2-bundle over S^2, sum stabilizations. For convenience, we let

$$\mathbb{S}_k := \begin{cases} S^2 \times S^2 & \text{for } k \text{ even,} \\ \mathbb{C}P^2 \# \overline{\mathbb{C}P^2} & \text{for } k \text{ odd,} \end{cases}$$

and note that $\mathbb{S}_k$ is described by the $(0, k)$-framed Hopf link.

Lemma 5.1 (blow-ups and sum stabilizations). *Let $\mathfrak{S} = (\Sigma, \Gamma)$ be a surface diagram and let $\mathfrak{S}'$ be obtained from $\mathfrak{S}$ by a substitution of type*

$$\big(a, b \mid a, \tau_b^{\pm 1}(a), b\big). \tag{5-1}$$

Furthermore, let $\mathfrak{S}''$ be obtained by a substitution of type

$$\big(a, b \mid a, b, \tau_b^{k}(a), b\big). \tag{5-2}$$

Then $Z_{\mathfrak{S}'}$ is diffeomorphic to the blow-up $Z_{\mathfrak{S}} \# \mp\mathbb{C}P^2$ and $Z_{\mathfrak{S}''}$ is diffeomorphic to the sum stabilization $Z_{\mathfrak{S}} \# \mathbb{S}_{-k}$.

Of course, any substitution is reversible so that whenever a surface diagram contains a configuration of the form $(a, \tau_b^{\pm 1}(a), b)$ or $(a, b, \tau_b^k(a), b)$ the associated 4-manifold must be a blow-up or sum stabilization, respectively. We will call these *blow-up* (respectively *sum stabilization*) *configurations*.

Proof. By switching we can assume that $\Gamma = (\dots, a, b)$, so $\Gamma'(\dots, a, \tau_b^{\pm 1}(a), b)$ and $\Gamma'' = (\dots, a, b, \tau_b^k(a), b)$. Figure 17 shows the relevant parts of the handle decompositions of the associated 4-manifolds. The shaded ribbons indicate the regions that contain all the other fold handles. Note that the curves a and b appear as 0-framed meridians to the dotted circles.

In the case of $Z_{\mathfrak{S}'}$ we can use the meridians to unlink the curve corresponding to $\tau_b^{\pm}(a)$ resulting in an unknot with framing ∓ 1 which is isolated from the rest of the diagram. Furthermore, the rest of the diagram agrees with the diagram for $Z_{\mathfrak{S}}$ and the claim follows.

The argument for $Z_{\mathfrak{S}''}$ is almost the same. Again, by sliding over the meridians we can isolate the curves corresponding to b and $\tau_b^k(a)$ from the rest of the diagram. This time we obtain a $(0, -k)$-framed Hopf link that represents a copy of $\mathbb{S}_{-k}$. $\square$

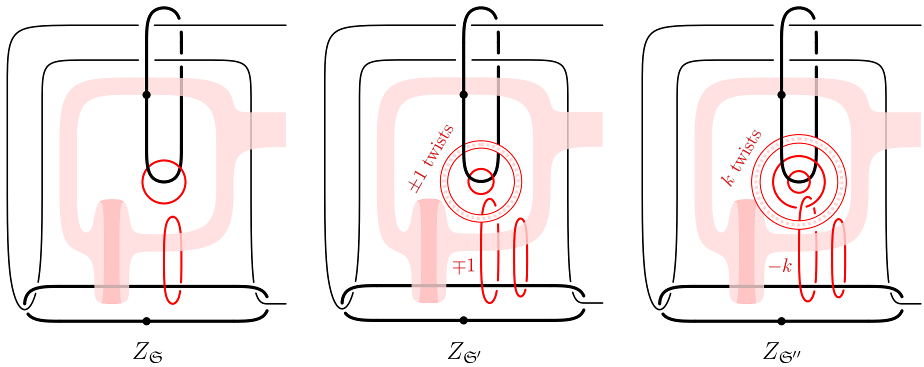

Figure 17. The relevant parts of the handle decompositions of $Z_{\mathfrak{S}}$, $Z_{\mathfrak{S}'}$ and $Z_{\mathfrak{S}''}$. All 2-handles without framing coefficient are 0-framed.

Proposition 5.2. *Let $\mathfrak{S}$, $\mathfrak{S}'$ and $\mathfrak{S}''$ be as in Lemma 5.1.*

(1) *All three diagrams have the same monodromy.*

(2) *If $\mathfrak{S}$ has trivial monodromy so that $Z_{\mathfrak{S}}$ closes off to a closed 4-manifold X, then $Z_{\mathfrak{S}'}$ closes off to $X \# \mp\mathbb{C}P^2$ and $Z_{\mathfrak{S}''}$ closes off to $X \# \mathbb{S}_k$.*

(3) *Any closed 4-manifold obtained from $\mathfrak{S}'$ (resp. $\mathfrak{S}''$) is a blow-up (resp. sum stabilization) of a manifold obtained from $\mathfrak{S}$.*

Proof. The first statement follows directly from Lemma 5.1 since connected sums with closed manifold (taken in the interior) do not change the boundary.

For the other statements, observe that if one knows how to apply the method from Remark 4.12 for $\mathfrak{S}$, then one also knows it for $\mathfrak{S}'$ and $\mathfrak{S}''$, and vice versa. □

Another instance where a substitution corresponds to a well known cut-and-paste operation was observed in [Hayano 2012, Lemma 6.13]. Assume that a surface diagram $\mathfrak{S}$ contains a curve $c \subset \Sigma$. If $d \subset \Sigma$ is geometrically dual to c, then one can perform a substitution of type $(c \mid c, d, c)$ and Hayano shows that if $\mathfrak{S}'$ denotes the resulting surface diagram, then $Z_{\mathfrak{S}'}$ is obtained from $Z_{\mathfrak{S}}$ by a surgery on the curve $\delta \subset \Sigma \subset Z_{\mathfrak{S}}$ with respect to its *fiber framing*, that is, the framing induced by the its canonical framing in Σ together with the framing of Σ in $Z_{\mathfrak{S}}$ as a regular fiber of $w_{\mathfrak{S}} : Z_{\mathfrak{S}} \to D^2$.

One immediately notices that our sum-stabilization substitution is a special case of this construction. However, it also paves the way for the following minor generalization of the surgery substitution which captures not only the fiber framed surgery but also the one with the opposite framing. (Recall that an embedded circle in an orientable 4-manifold always has trivial normal bundle and there are exactly two framings, since $\pi_1(\mathrm{SO}(3)) \cong \mathbb{Z}_2$.)

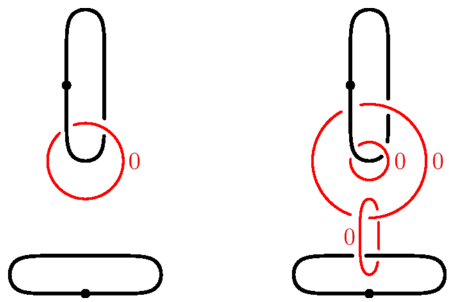

Figure 18. Hayano's surgery substitution: neighborhoods with vanishing cycle c (left) and vanishing cycles c, d, c (right).

Lemma 5.3. *Let $\mathfrak{S}$ and $\mathfrak{S}'$ be two surface diagrams with the same underlying surface Σ and let $c, d \subset \Sigma$ be a geometrically dual pair of simple closed curves. If $\mathfrak{S}'$ is obtained from $\mathfrak{S}$ by a substitution of type $(c \mid c, \tau_c^k(d), c)$, then $Z_{\mathfrak{S}'}$ is obtained from $Z_{\mathfrak{S}}$ by a surgery on $d \subset \Sigma \subset X$ with respect to the fiber framing when k is even and the opposite framing when k is odd.*

Proof. As in Hayano's proof, it is enough to work in a neighborhood of $c \cup d$ which we can assume to be a punctured torus. Using our handle decomposition instead of the ones from broken Lefschetz fibrations, the effect of Hayano's surgery substitution, that is, the case when $k = 0$, looks as in Figure 18, where c (respectively d) appears as the meridian of the upper (respectively lower) 1-handle. To obtain the other even cases, observe that in Figure 18, right, we can slide the 2-handle corresponding to d once over each 2-handle corresponding to c in the same direction. Depending on the direction this changes the framing coefficient by ± 2 and one readily checks that the resulting diagram shows a neighborhood with vanishing cycles $(c, \tau_c^{\mp 2}(d), c)$. Repeating this trick one can obtain all configurations with even k and they will all describe the fiber framed surgery on d.

As shown in [Gompf and Stipsicz 1999, Example 8.4.6] the surgery with the opposite framing can be realized by inserting a pair of a Lefschetz vanishing cycle and an achiral Lefschetz vanishing cycle which are both parallel to d. But Figure 19 shows that the result is the same as a substitution of type $(c \mid c, \tau_c^{-1}(d), c)$ which corresponds to $k = -1$. Moreover, the arguments for shifting the value of k by multiples of 2 works just as in the fiber framed case. $\square$

Using Lemma 5.3, the sum stabilization can be interpreted as performing surgery on a null-homotopic curve with either of its framing. Indeed, as d one takes one of the adjacent vanishing cycles of c in $\mathfrak{S}$ which is clearly null-homotopic in $Z_{\mathfrak{S}}$.

It would be interesting to interpret other cut-and-paste operations on 4-manifolds as substitutions in surface diagrams. For example, it is reasonable to expect such

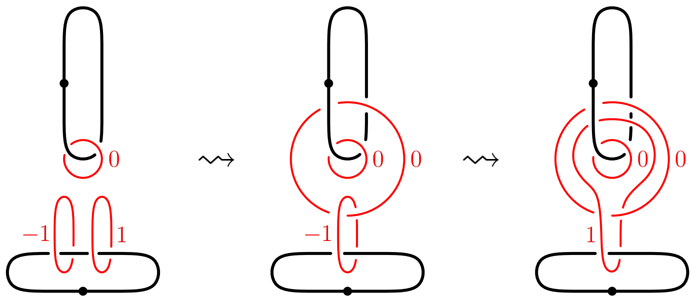

Figure 19. Surgery with the opposite framing.

an interpretation for certain rational blow downs which can be described in terms of Lefschetz fibrations; see [Endo et al. 2011]. However, we settle for blow-ups and sum stabilizations in this paper.

6. Manifolds with genus-one simple wrinkled fibrations

In this section we prove Theorem 1.2. Our strategy is to use Proposition 5.2 to construct some genus-one simple wrinkled fibrations and then show that this construction gives all such fibrations.

We begin with the construction of genus-one simple wrinkled fibrations over S^2. As before, we denote by $\mathbb{S}_k$ the closed 4-manifolds described by the $(0, k)$-framed Hopf link and we define a family of manifolds

$$X_{klmn} = \mathbb{S}_k \# l(S^2 \times S^2) \# m\mathbb{C}P^2 \# n\overline{\mathbb{C}P^2}, \quad k \in \{0, 1\}, \; l, m, n \geq 0. \tag{6-1}$$

Note that these are precisely the manifolds in Theorem 1.2. Recall that $\mathbb{S}_k$ is an S^2-bundle over S^2. By performing a birth on a suitable bundle projection $\mathbb{S}_k \to S^2$ we obtain a simple wrinkled fibration with two cusps. We can then use Proposition 5.2 to add the other summands at will. Thus, in order to prove Theorem 1.2, it remains to show the following.

Proposition 6.1. *Let $w : X \to S^2$ be a simple wrinkled fibration of genus one. Then X is diffeomorphic to some X_{klmn} described in* (6-1).

Remark 6.2. The reason for our small reformulation of Theorem 1.2 is that, while the original formulation is cleaner, the new one is more in tune with the structure of the proof.

The key to the proof of Proposition 6.1 is the simple nature of simple closed curves on the torus. Indeed, the two facts that two oriented simple closed curves on the torus are isotopic if and only if they are homologous and that the (absolute value of the) algebraic and geometric intersection numbers agree allow us to transfer the whole discussion of genus 1 surface diagrams into the homology group $H_1(T^2) \cong \mathbb{Z} \oplus \mathbb{Z}$ simply by choosing orientations on the curves. Building on this observation we obtain the following result about the structure of genus-one surface diagrams.

Lemma 6.3. *Any closed circuit on the torus of length at least three contains blow-up or sum stabilization configurations* (*as described in Lemma 5.1*).

Proof. Let $\Gamma = (c_1, \dots, c_l)$ be a (not necessarily closed) circuit on the torus of length $c \geq 3$. As usual, we choose an arbitrary orientation on c_1 and orient the remaining curves by requiring that $\langle c_i, c_{i+1} \rangle = 1$ for $i < l$ so that we can consider each c_i as an element of $H_1(T^2)$.

We first observe that, since any two adjacent curves in Γ are algebraically dual, they form a basis of $H_1(T^2)$. In particular, for $i \geq 3$ we can write

$$c_i = k_i c_{i-1} - c_{i-2}, \quad k_i \in \mathbb{Z}$$

where the coefficient of c_{i-2} determined by our convention that $\langle c_{i-1}, c_i \rangle = 1$. This shows that if we denote by $\sigma_i := \langle c_1, c_i \rangle$ the algebraic intersection number between c_1 and c_i, then we obtain a recursive formula

$$\sigma_i = k_i \sigma_{i-1} - \sigma_{i-2} \tag{6-2}$$

for $i \geq 3$ with initial values $\sigma_1 = 0$ and $\sigma_2 = 1$. At this point we note that Γ is closed if and only if $|\sigma_l| = 1$.

We claim that if $|k_i| \geq 2$ for all $i \geq 3$, then $|\sigma_{i+1}| > |\sigma_i|$ for all i. This follows inductively since $|\sigma_2| > |\sigma_1|$ and from (6-2) we get

$$|\sigma_{i+1}| = |k_{i+1}\sigma_i - \sigma_{i-1}| \geq \big| |k_{i+1}||\sigma_i| - |\sigma_{i-1}| \big| = |k_{i+1}||\sigma_i| - |\sigma_{i-1}| > |\sigma_i|,$$

where we have used the reverse triangle inequality, the induction hypothesis and the assumption that $|k_{i+1}| \geq 2$. As a consequence, we see that if Γ is closed, then we must have $|k_i| \leq 1$ for some $i \geq 3$.

Assume first that $k_i = \pm 1$. To keep the notation transparent we momentarily rename the relevant curves to

$$(c_{i-2}, c_{i-1}, c_i) =: (a, \xi, b). \tag{6-3}$$

By assumption, $b = \pm\xi - a$ and thus $\xi = \pm(a + b)$ and the orientation convention shows that $\langle a, b \rangle = \pm 1$. By invoking the Picard–Lefschetz formula (Proposition 2.7) we obtain

$$\tau_a^{\pm 1}(b) = b \pm \langle a, b \rangle\, a = a + b = \pm\xi,$$

which, after forgetting the orientations again, reveals the excerpt of Γ shown in (6-3) as a blow-up configuration.

A similar argument exhibits a sum-stabilization configuration in the remaining case when $k_i = 0$. The details are left to the reader. $\square$

The proof of Proposition 6.1, and thus of Theorem 1.2, is now very easy.

Proof of Proposition 6.1. Any genus-one simple wrinkled fibration over S^2 is obtained by closing off a manifold $Z_{\mathfrak{S}}$ associated to a surface diagram $\mathfrak{S} = (T^2, \Gamma)$. Moreover, any such diagram $\mathfrak{S}$ can be closed off since the mapping class group of the lower-genus fiber is trivial. By Lemma 6.3 and Proposition 5.2(3) we can successively split off summands of the form $\pm\mathbb{C}P^2$ and $\mathbb{S}_k$ until the remaining surface diagram, say $\mathfrak{S}_0$ has a circuit of length two. It is easy to see that $Z_{\mathfrak{S}_0}$ is the trivial disk bundle $S^2 \times D^2$. (Either by drawing a Kirby diagram or by observing that any simple wrinkled fibration with two cusps is homotopic to a bundle projection.) Thus there are exactly two ways to close off the fibration, producing a summand of the form $\mathbb{S}_0 \cong S^2 \times S^2$ or $\mathbb{S}_1 \cong \mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$. □

7. Concluding remarks

The theory of simple wrinkled fibrations and surface diagrams is still in a very early stage and at this point it raises more questions then it provides answers. We would like to point out what we consider as some of the major problems in the subject as well as to indicate some further developments.

7A. ***Closed 4-manifolds.*** The ultimate goal is to use surface diagrams to study *closed* 4-manifolds. Unfortunately, it turns out that most surface diagrams do *not* describe closed manifolds since they have nontrivial monodromy and it is usually a hard problem to determine whether a given surface diagram has trivial monodromy. The following is thus of great interest.

Problem 7.1. Find at least necessary conditions for a surface diagram to have trivial monodromy that are easier to check.

The next major problem was already mentioned on page 292. If a surface diagram of sufficiently high genus is known to have trivial monodromy, then it determines a unique closed 4-manifold together with a simple wrinkled fibration over S^2 by closing off the associated fibration over the disk. However, for practical purposes the information on how to close off is encoded too implicitly in the surface diagram. For example, by simply looking at the surface diagram it not at all clear how to answer the following very reasonable questions about the corresponding simple wrinkled fibration over S^2:

- Does the fibration have a section?
- What can be said about the homology class of the fiber? Is it trivial, primitive, torsion, ... ?
- What is the fundamental group, homology, etc. of the total space?

What is missing is one more piece of information which is roughly the (framed) attaching curve of the last 2-handle. One can also reformulate this issue in terms of mapping class groups (see [Hayano 2012], for example).

Problem 7.2. Find a practical method to determine the missing piece of information from a surface diagram with trivial monodromy.

7B. ***Higher-genus fibrations.*** The fact that any (achiral) Lefschetz fibration can be turned into a simple wrinkled fibration of one genus higher suggests the philosophy that simple wrinkled fibrations of a fixed genus might behave similarly as (achiral) Lefschetz fibrations of one genus lower.

This analogy works rather well for the lowest possible fiber genera. Indeed, our result about genus-one simple wrinkled fibrations looks very similar to the (rather trivial) classification of genus zero (achiral) Lefschetz fibrations, the latter being blow-ups of either $S^2 \times S^2$ or $\mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$.

Following this train of thought one might hope to be able to say something useful about the classification of genus two simple wrinkled fibrations over S^2 but one should expect to be lost as soon as the genus is three or higher. However, it is nonetheless conceivable that part of the classification scheme that works in the genus-one case might carry over to higher-genus fibrations, as we will now explain.

Let $\mathfrak{S} = (\Sigma; c_1, \dots, c_l)$ be a surface diagram and assume that for some $2 < k < l$ the curve c_k is geometrically dual to c_1. Then there is an obvious way to decompose $\mathfrak{S}$ into the two smaller surface diagrams $(\Sigma; c_1, \dots, c_k)$ and $(\Sigma; c_1, c_k, \dots, c_l)$. Repeating this process we eventually obtain a decomposition of $\mathfrak{S}$ into a collection of surface diagrams with the property that no pair of nonadjacent curves has geometric intersection number one. Let us call such a surface diagram *irreducible*.

In terms of the simple wrinkled fibration associated to $\mathfrak{S}$ the above decomposition of $\mathfrak{S}$ should correspond to merging the fold arcs that induce c_1 and c_k. (As shown in [Lekili 2009], the necessary and sufficient condition for a fold merge is exactly that the vanishing cycles of the fold arcs are geometrically dual.) The result is a wrinkled fibration that naturally decomposes as a *boundary fiber sum* of the two simple wrinkled fibrations associated to the parts of the decomposition of $\mathfrak{S}$.

This suggests that any descending simple wrinkled fibration over the disk naturally decomposes into a boundary fiber sum of *irreducible* fibrations where we call a simple wrinkled fibration irreducible if its surface diagram is irreducible. Consequently, the classification of descending simple wrinkled fibrations splits into two parts: the classification of irreducible fibrations and understanding the effect of boundary fiber sums.

The genus-one classification fits into this scheme as follows. Our arguments show that the only irreducible surface diagrams of genus-one are given by the blow-up configurations $(a, \tau_a^{\pm 1}(b), b)$ and the sum-stabilization configurations

$(a, b, \tau_b^k(a), b)$ for $k \neq 1$. Using the handle decompositions it is easy to identify the corresponding manifolds. (They are the connected sum of $S^2 \times D^2$ with either $\pm\mathbb{C}P^2$, $S^2 \times S^2$ or $\mathbb{C}P^2 \# \overline{\mathbb{C}P^2}$.) Furthermore, the boundary fiber sums are performed along spheres and are thus easy to understand.

Making these arguments precise requires an understanding of the effect of merging folds and cusps on surface diagrams.

7C. *Uniqueness of surface diagrams.* Given the fact that all closed 4-manifolds can be described by surface diagrams, it is natural to ask for a set of moves to relate different surface diagrams that describe the same manifold, similar to the situation of 3-manifolds and Heegaard diagrams.

A first step in this direction was taken by Williams [2011] who relates the surface diagrams of homotopic simple wrinkled fibrations over S^2 of genus at least three. He shows that any two homotopic simple wrinkled fibrations can be connected by a special homotopy that is made up of four basic building blocks. These building blocks are simple enough to understand their effect on the initial surface diagram (see also [Hayano 2012]).

So far this is completely analogous to the 3-dimensional context. A new phenomenon in the 4-dimensional context is that two simple wrinkled fibrations on a given 4-manifold are not necessarily homotopic. The structure of the set $\pi^2(X) := [X, S^2]$ of homotopy classes of maps from a closed 4-manifold to the 2-sphere — also known as the *second cohomotopy set* of X — is described in [Kirby et al. 2012] (see also the references therein). Our results show that an equivalence class of surface diagrams for X determines an orbit of the action of the diffeomorphism group of X on $\pi^2(X)$. This action is usually neither trivial, as shown by the two projections of $S^2 \times S^2$ which are interchanged by flipping the factors, nor transitive since the action of the diffeomorphism group on the second homology group preserves divisibility. Thus, reparametrizing a surface diagram can change the homotopy class of its simple wrinkled fibration but one cannot expect to obtain all homotopy classes in this way.

A general method for relating broken fibrations in different homotopy classes is the *projection move* mentioned in [Williams 2010] but it is not at all obvious how to interpret this procedure in terms of surface diagrams. Altogether, the problem of relating surface diagram with nonhomotopic fibrations is still wide open.

Acknowledgements

This work is part of the author's ongoing PhD project carried out at the Max Planck Institute for Mathematics in Bonn, Germany. The author would like to thank his advisor Prof. Dr. Peter Teichner for his support as well as İnanç Baykur for helpful comments on an early draft of this paper. The author is supported by an IMPRS Scholarship of the Max Planck Society.

References

[Akbulut and Karakurt 2008] S. Akbulut and Ç. Karakurt, “Every 4-manifold is BLF”, *J. Gökova Geom. Topol. GGT* **2** (2008), 83–106. MR 2009k:57037 Zbl 1209.57015

[Auroux et al. 2005] D. Auroux, S. K. Donaldson, and L. Katzarkov, “Singular Lefschetz pencils”, *Geom. Topol.* **9** (2005), 1043–1114. MR 2005m:53164 Zbl 1077.53069

[Baykur 2008] R. İ. Baykur, “Existence of broken Lefschetz fibrations”, *Int. Math. Res. Not.* **2008** (2008), Art. ID rnn 101, 15. MR 2010b:57026 Zbl 1179.57011

[Baykur 2009] R. İ. Baykur, “Topology of broken Lefschetz fibrations and near-symplectic four-manifolds”, *Pacific J. Math.* **240**:2 (2009), 201–230. MR 2010c:57035 Zbl 1162.57011

[Baykur and Kamada 2010] R. I. Baykur and S. Kamada, “Classification of broken Lefschetz fibrations with small fiber genera”, preprint, 2010. arXiv 1010.5814v2

[Boardman 1967] J. M. Boardman, “Singularities of differentiable maps”, *Inst. Hautes Études Sci. Publ. Math.* 33 (1967), 21–57. MR 37 #6945 Zbl 0165.56803

[Donaldson 1999] S. K. Donaldson, “Lefschetz pencils on symplectic manifolds”, *J. Differential Geom.* **53**:2 (1999), 205–236. MR 2002g:53154 Zbl 1040.53094

[Earle and Eells 1969] C. J. Earle and J. Eells, “A fibre bundle description of Teichmüller theory”, *J. Differential Geometry* **3** (1969), 19–43. MR 43 #2737a Zbl 0185.32901

[Endo and Gurtas 2010] H. Endo and Y. Z. Gurtas, “Lantern relations and rational blowdowns”, *Proc. Amer. Math. Soc.* **138**:3 (2010), 1131–1142. MR 2011a:57047 Zbl 1198.57017

[Endo et al. 2011] H. Endo, T. E. Mark, and J. Van Horn-Morris, “Monodromy substitutions and rational blowdowns”, *J. Topol.* **4**:1 (2011), 227–253. MR 2012b:57051 Zbl 1234.57037

[Farb and Margalit 2011] B. Farb and D. Margalit, *A primer on mapping class groups*, Princeton Mathematical Series **49**, Princeton University Press, 2011. MR 2012h:57032 Zbl 1245.57002

[Gay and Kirby 2007] D. T. Gay and R. Kirby, “Constructing Lefschetz-type fibrations on four-manifolds”, *Geom. Topol.* **11** (2007), 2075–2115. MR 2009b:57048 Zbl 1135.57009

[Gay and Kirby 2011a] D. T. Gay and R. Kirby, “Indefinite Morse 2-functions; broken fibrations and generalizations”, preprint, 2011. arXiv 1102.0750v2

[Gay and Kirby 2011b] D. T. Gay and R. C. Kirby, “Fiber-connected, indefinite Morse 2-functions on connected n-manifolds”, *Proc. Natl. Acad. Sci. USA* **108**:20 (2011), 8122–8125. MR 2012k:57041

[Gay and Kirby 2012] D. T. Gay and R. Kirby, “Reconstructing 4-manifolds from Morse 2-functions”, preprint, 2012. Zbl 06124834 arXiv 1202.3487

[Golubitsky and Guillemin 1973] M. Golubitsky and V. Guillemin, *Stable mappings and their singularities*, Graduate Texts in Mathematics **14**, Springer, New York, 1973. MR 49 #6269 Zbl 0294.58004

[Gompf and Stipsicz 1999] R. E. Gompf and A. I. Stipsicz, 4-*manifolds and Kirby calculus*, Graduate Studies in Mathematics **20**, American Mathematical Society, Providence, RI, 1999. MR 2000h:57038 Zbl 0933.57020

[Hayano 2011] K. Hayano, “On genus-1 simplified broken Lefschetz fibrations”, *Algebr. Geom. Topol.* **11**:3 (2011), 1267–1322. MR 2012f:57067 Zbl 1229.57017

[Hayano 2012] K. Hayano, “Modification rule of monodromies in R_2-move”, preprint, 2012. arXiv 1203.4299v1

[Ivanov 1992] N. V. Ivanov, *Subgroups of Teichmüller modular groups*, Translations of Mathematical Monographs **115**, American Mathematical Society, Providence, RI, 1992. MR 93k:57031 Zbl 0776.57001

[Kas 1980] A. Kas, "On the handlebody decomposition associated to a Lefschetz fibration", *Pacific J. Math.* **89**:1 (1980), 89–104. MR 82f:57012 Zbl 0457.14011

[Kirby et al. 2012] R. Kirby, P. Melvin, and P. Teichner, "Cohomotopy sets of 4-manifolds", preprint, 2012. arXiv 1203.1608v1

[Lekili 2009] Y. Lekili, "Wrinkled fibrations on near-symplectic manifolds", *Geom. Topol.* **13**:1 (2009), 277–318. MR 2009k:57043 Zbl 1164.57006

[Levine 1964] H. I. Levine, "The singularities, ${S_1}^q$", *Illinois J. Math.* **8** (1964), 152–168. MR 28 #2560 Zbl 0124.38801

[Morin 1965] B. Morin, "Formes canoniques des singularités d'une application différentiable", *C. R. Acad. Sci. Paris* **260** (1965), 5662–5665. MR 31 #5212 Zbl 0178.26801

[Williams 2010] J. Williams, "The h-principle for broken Lefschetz fibrations", *Geom. Topol.* **14**:2 (2010), 1015–1061. MR 2011d:57066 Zbl 1204.57027

[Williams 2011] J. D. Williams, "Topology of surface diagrams of smooth 4-manifolds", *Proc. Natl. Acad. Sci. USA* **108**:20 (2011), 8126–8130. MR 2012f:57066

Received August 3, 2012. Revised October 12, 2012.

STEFAN BEHRENS
MAX PLANCK INSTITUTE FOR MATHEMATICS
VIVATSGASSE 7
D-53111 BONN
GERMANY

sbehrens@mpim-bonn.mpg.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.307

THIN r-NEIGHBORHOODS OF EMBEDDED GEODESICS WITH FINITE LENGTH AND NEGATIVE JACOBI OPERATOR ARE STRONGLY CONVEX

PHILIPPE DELANOË

In a complete Riemannian manifold, an embedded geodesic γ with finite length and negative Jacobi operator admits an r-neighborhood $N_r(\gamma)$ with radius $r > 0$ small enough such that each pair of points of $N_r(\gamma)$ can be joined by a unique geodesic contained in $N_r(\gamma)$ where it minimizes length among the piecewise C^1 paths joining its endpoints.

Introduction

Let M be a connected complete Riemannian manifold; let d denote its Riemannian distance function [do Carmo 1992]. A connected subset $S \subset M$ with nonempty interior S° is called strongly convex for a pair of points $(p, q) \in S \times S$ if there exists a unique geodesic path $t \in [0, 1] \to \gamma(t) \in M$ such that $\gamma(0) = p$, $\gamma(1) = q$ and $\gamma(t) \in S^\circ$ for $t \in (0, 1)$, with γ length-minimizing among piecewise C^1 paths from p to q in $\bar{S}$. The subset S is just called *strongly convex* if it is so for each pair $(p, q) \in S \times S$.

Definition 0.1. Let $S \subset M$ be a strongly convex subset. For each pair $(p, q) \in S \times S$, the length of the geodesic path joining p to q with interior in S° is called the *inner distance* from p to q in S, denoted by $d_S(p, q)$.

It is quite natural to endow a strongly convex subset $S \subset M$ with its inner distance function d_S. The latter is nothing but the length metric associated with the metric space $(S, d|_S)$ [Gromov 1981].

Since Whitehead's landmark paper [1932], it has been known that small enough balls in M are strongly convex. Moreover, if B is such a ball, its inner distance function d_B coincides with the restriction of d to $B \times B$ [Kobayashi and Nomizu 1996; Cheeger and Ebin 2008; Aubin 1998; do Carmo 1992; Klingenberg 1995]. In the flat torus $\mathbb{R}^n/\mathbb{Z}^n$, if the radius of a ball B belongs to the interval $\left(\frac{1}{4}, \frac{1}{2}\right)$, the reader can check that B remains strongly convex but d_B no longer coincides

Supported by the CNRS at UNS.
MSC2000: primary 34C11, 53C21, 53C22; secondary 53B21.
Keywords: geodesic, negative curvature, Fermi chart, tube, convexity.

with $d|_{B\times B}$. Here, we would like to construct a general family of examples of strongly convex subsets $S \subset M$ such that $d_S \not\equiv d|_{S\times S}$.

The notion of *extended* distance function used in [Figalli et al. 2012] is similar in spirit to that of inner metric; could it guide us toward an example? Let us recall its definition. If $t \in [0, 1] \to \gamma(t) \in M$ is an embedded geodesic without conjugate points, the map $\mathrm{Id} \times \exp : TM \to M \times M$ induces a diffeomorphism Ψ_γ from a neighborhood $\mathcal{U}$ of $\big(\gamma(0), (d\gamma/dt)(0)\big)$ in TM to a neighborhood $\mathcal{W}$ of $(\gamma(0), \gamma(1))$ in $M \times M$. The extended distance function d_γ of [Figalli et al. 2012] is then defined in $\mathcal{W}$ by $d_\gamma(p, q) = |V|_p$ where $\Psi_\gamma(p, V) = (p, q)$. It is called so because, if γ contains no cut point, shrinking $\mathcal{W}$ if necessary, it satisfies $d_\gamma(p, q) \equiv d(p, q)$. In this setting, we would like to know whether a thin enough tube about the geodesic γ must be strongly convex. Anytime it is, one may identify d_γ with the restriction to $\mathcal{W}$ of the inner distance function of the tube; in particular, the function d_γ satisfies in effect the distance axioms.

By a *tube* about γ is meant a closed subset of M containing $\gamma([0, 1])$, with nonempty interior and each point of which admitting a unique *nearest point* in $\gamma([0, 1])$; moreover, if $p \mapsto p^\perp_\gamma$ denotes the nearest-point map, the geodesic from p to $p^\perp_\gamma$ should meet $\gamma([0, 1])$ *orthogonally*. Finally, the lateral boundary of the tube is given by the equation $d(p, p^\perp_\gamma) = r$, where $r > 0$ is a small real number called the *radius* of the tube.

We are thus willing to study the question: *under which conditions must a tube about an embedded geodesic be strongly convex*?

First of all, indeed, we should restrict to geodesics without conjugate points (at least in their interior) since, by the Morse index theorem, they would not be minimizing otherwise [Milnor 1963]. To proceed further, let us take examples. In the domain of the unit sphere of $\mathbb{R}^3$ given by $0 \leqslant$ longitude $< \pi$ and $-r \leqslant$ latitude $\leqslant r$ with r small, we see that the geodesic joining two points with equal latitude close enough to r does not stay in that domain. But if we look at a similar domain about the *interior* equator of a torus of revolution in $\mathbb{R}^3$ and pick two points as above, the geodesic joining them does stay in the domain. So, a curvature assumption should be made along a geodesic before we can expect the strong convexity of a tube about it, and positive curvature rules out strong convexity.

Eventually, we will show that a tube $T_r(\gamma_0)$ with small enough radius r about a geodesic γ_0 with negative Jacobi operator is essentially strongly convex. Specifically, we will prove the following result:

Theorem 0.2. *Let $\gamma_0 : s \in [0, \ell_0] \to \gamma_0(s) \in M$ be an embedded unit-speed geodesic with negative Jacobi operator. Given $\varsigma > 0$, there exists $\varrho > 0$ such that, if $r \in (0, \varrho)$, the tube $T_r(\gamma_0)$ is strongly convex for each pair $(p, q) \in T_r(\gamma_0) \times T_r(\gamma_0)$ of points satisfying either $|s(p^\perp_{\gamma_0}) - s(q^\perp_{\gamma_0})| \geqslant \varsigma$, or $s(p^\perp_{\gamma_0})$ and $s(q^\perp_{\gamma_0})$ belong to the subinterval*

$[\varsigma, \ell_0 - \varsigma]$. *Furthermore, if M has dimension* 2, *the result holds with* $\varsigma = 0$ *provided we except the boundary pairs* (p, q) *lying in the same end* ($s = 0$ *or* $s = \ell_0$) *of the tube.*

In this statement, we allow the geodesic γ_0 to contain cut points. For instance, if the image of γ_0 is contained in the curve $\{x^2 + y^2 = 1, z = 0\}$ viewed as the interior equator of a torus of revolution in $\mathbb{R}^3$, we allow its length ℓ_0 to belong to the interval $[0, 2\pi)$. In this context, the inner distance function for which we are looking appears well approximated by the pseudometric defined in the tube by $\hat{d}(p,q) = |s(p^{\perp}_{\gamma_0}) - s(q^{\perp}_{\gamma_0})|$, at least for the pairs $(p,q) \in T_r(\gamma_0) \times T_r(\gamma_0)$ such that $\hat{d}(p,q) \gg r$. Accordingly, our proof will split in two parts; let us provide a rough outline of it.

<u>Case 1</u>: For $\hat{d}(p,q)$ less than a suitable positive constant c independent of r as $r \downarrow 0$, there exists a unique minimizing geodesic $t \in [0,1] \to \gamma(t) \in M$ from p to q, so we only have to prove the inclusion $\gamma((0,1)) \subset (T_r(\gamma_0))^{\circ}$. We do it using a one-parameter family of geodesics $\lambda \in [0,1] \to c_\lambda$ interpolating between c_0 given by $t \in [0,1] \to \gamma_0\big(ts(q^{\perp}_{\gamma_0}) + (1-t)s(p^{\perp}_{\gamma_0})\big)$ and $c_1 = \gamma$. For λ small, we certainly have $c_\lambda((0,1)) \subset (T_r(\gamma_0))^{\circ}$. We must rule out the possibility that $c_\lambda(t)$ first touches the boundary of $T_r(\gamma_0)$ for some $t \in (0,1)$. If $n = 2$, it could happen but on the *lateral* part of $\partial T_r(\gamma_0)$ because the ends of $T_r(\gamma_0)$ are *totally geodesic*. If $n > 2$, the pinching $s[(c_\lambda(t))^{\perp}_{\gamma_0}] \in (0, \ell_0)$ is obtained relying on the assumption (ignored elsewhere in the proof) that $\hat{d}(p,q) \geqslant \varsigma$ or $s(p^{\perp}_{\gamma_0})$ and $s(q^{\perp}_{\gamma_0})$ lie in $[\varsigma, \ell_0 - \varsigma]$. As for the lateral part of $T_r(\gamma_0)$, the estimate $d\big(c_\lambda(t), (c_\lambda(t))^{\perp}_{\gamma_0}\big) < r$ (unless $p = q$) follows from a *maximum principle* for geodesics shown to hold in $T_r(\gamma_0)$ due to our curvature assumption.

<u>Case 2</u>: $\hat{d}(p,q) \geqslant c$. Here, we must work harder, shrink $r > 0$ and show that, if $t \in [0,1] \to \gamma(t) \in M$ is a geodesic from p to q ranging in $T_r(\gamma_0)$, its Jacobi operator should stay, like the one of γ_0, *negative*. Moreover, we infer from the latter property that γ must be minimizing and unique. We are thus left with proving the very *existence* of γ. It will be done by a tricky connectedness argument, fixing p, letting q vary in the tube and using the parameter $z = \hat{d}(p,q) \in [c, \ell_0]$ itself. The openness part of that argument is based on the invertibility of $d(\exp_p)(\dot{\gamma}(0))$, which holds due to the curvature property of γ; the closedness part relies on the aforementioned maximum principle.

Can one find a quicker proof? We did not. With Theorem 0.2 and its proof at hand, it becomes easy to obtain a full strong convexity result if, instead of the tube $T_r(\gamma_0)$, we consider the closure of the r-neighborhood of γ_0, that is, the subset $N_r(\gamma_0) = \{m \in M, d(\gamma_0([0,\ell_0]), m) \leqslant r\}$. In this way, we get the main result of the paper, namely:

Corollary 0.3 (main result). *Let* $\gamma_0 : s \in [0, \ell_0] \to \gamma_0(s) \in M$ *be an embedded unit-speed geodesic with negative Jacobi operator. There exists* $\varrho > 0$ *such that the subset* $N_r(\gamma_0) \subset M$ *is strongly convex for* $r \in (0, \varrho)$.

The paper is organized as follows: the next two sections are devoted to preliminary tools for the proof, general properties of thin tubes are recorded in Section 1 and further ones under our curvature assumption in Section 2, the proof of Theorem 0.2 itself is given in Section 3, and that of Corollary 0.3, in Section 4.

1. Properties of a thin tube about an embedded geodesic

Throughout this section, we use the setting of Theorem 0.2 but drop the assumption made on the Jacobi operator of the geodesic γ_0.

1A. ***Fermi map, cylinders and Gauss lemma.*** Let us recall how the tube $T_r(\gamma_0)$ can be precisely defined [Aubin 1998; Gray 2004]. The geodesic γ_0 extends uniquely as a geodesic embedding of an interval $I = (-\epsilon, \ell_0+\epsilon)$ with ϵ small. We consider the map

$$(V, s) \in V_0^{\perp} \times I \to E_0(V, s) = \exp^{\perp}_{\gamma_0(s)}(\|^{\gamma_0}(V)) \in M,$$

where we have denoted by $V_0^{\perp}$ the subspace of $T_{\gamma_0(0)}M$ orthogonal to the velocity vector $V_0 = (d\gamma_0/ds)(0)$, by $\|^{\gamma_0}(V)$ the vector field along γ_0 obtained by parallel transport of the vector V and by $\exp^{\perp}_{\gamma_0(s)}$ the restriction of the exponential map to $\|^{\gamma_0}(V_0)(s)^{\perp}$. The differential of E_0 at $(0, s)$ is given by

$$(\delta V, \delta s) \in V_0^{\perp} \times \mathbb{R} \to dE_0(0, s)(\delta V, \delta s) = \frac{d\gamma_0}{ds}(s)\delta s + \|^{\gamma_0}(\delta V)(s) \in T_{\gamma_0(s)}M;$$

it is an isomorphism since orthogonality is preserved by parallel transport. From the inverse function theorem [Lang 2002] and the compactness of $[0, \ell_0]$ (or bounded length of γ_0), we infer[1] the existence of a real $R > 0$ such that, setting $|V|$ for the norm of a vector V and $\overline{B}^{\perp}(0, R) = \{V \in V_0^{\perp}, |V| \leqslant R\}$, the map E_0 induces a *diffeomorphism* from a neighborhood of $\overline{B}^{\perp}(0, R) \times [0, \ell_0]$ onto a neighborhood of its image. Let us fix such a radius R once for all. For $r \leqslant R$, we denote by $T_r(\gamma_0)$ the image by E_0 of $\overline{B}^{\perp}(0, r) \times [0, \ell_0]$ and call it the *tube* about γ_0 with radius r [Gray 2004]. We set $p \mapsto F_0(p) = (v_0^{\perp}(p), z(p))$ for the *inverse* of the mapping E_0 and refer to it as the *Fermi map along* γ_0. We call $z(p)$ the *height* of the point p relative to γ_0 and the subsets $E_R^{\text{top}}(\gamma_0) = \{p \in T_R(\gamma_0), z(p) = \ell_0\}$ and $E_R^{\text{bot}}(\gamma_0) = \{p \in T_R(\gamma_0), z(p) = 0\}$, respectively, the *top and bottom ends* of the

[1] Full details are given in Section 1D for a construction encompassing the present one.

tube. If $p \in T_R(\gamma_0)$, the unit-speed geodesic

$$s \in [0, |v_0^\perp(p)|] \to E_0\left(s\frac{v_0^\perp(p)}{|v_0^\perp(p)|}, z(p)\right)$$

is the unique minimizing geodesic from γ_0 to p; its length $\mathfrak{r}_{\gamma_0}(p) = |v_0^\perp(p)|$ is thus equal to $d(\gamma_0, p)$. For short, that geodesic will be denoted by $s \mapsto [\gamma_0, p](s) \in T_R(\gamma_0)$, and the function $\mathfrak{r}_{\gamma_0}$ itself simply by $\mathfrak{r}$ unless a confusion may occur. We let $N_{\gamma_0}(p)$, or just $N(p)$ if no confusion, denote the velocity vector $d[\gamma_0, p]/ds$ evaluated at $s = d(\gamma_0, p)$. The unit vector field $p \mapsto N(p)$ is defined in the open subset of the tube $T_R(\gamma_0)$ where $\mathfrak{r}(p) > 0$, that is, outside the geodesic γ_0; moreover, it is readily seen to satisfy $dz(N) = 0$, $d\mathfrak{r}(N) = 1$ and $\nabla_N N = 0$, with ∇ the Levi-Civita connection. If $r \in (0, R]$, we set $C_r(\gamma_0) = \{p \in T_R(\gamma_0), \mathfrak{r}(p) = r\}$ for the *cylinder* of radius r about γ_0, sometimes called the *lateral* part of the boundary of the tube $T_r(\gamma_0)$. The outward unit *normal* to that cylinder at $p \in C_r(\gamma_0)$ is nothing but $N(p)$ due to the *generalized Gauss lemma* according to which the gradient of the function $\mathfrak{r}$ and the vector field N coincide [Gray 2004, pp. 26–28]. The identity $N = \operatorname{grad} \mathfrak{r}$ will be central for us. It yields the following identity, recorded here for later use, valid at each $p \in T_R(\gamma_0)$ such that $\mathfrak{r}(p) > 0$:

$$(1)\quad (g-d\mathfrak{r}^2)(V,W)=(g-d\mathfrak{r}^2)(\Pi_N^\perp(V),\Pi_N^\perp(W)) \quad \text{for all } (V,W)\in T_pM\times T_pM,$$

where we have set $\Pi_N^\perp(V) = V - g(V, N)N$ for the orthogonal projection of T_pM onto $N(p)^\perp$; in other words, if we write $TM = \mathbb{R}N \oplus N^\perp$ on $\{\mathfrak{r} > 0\}$, the generalized Gauss lemma implies that the metric g splits into the sum of $d\mathfrak{r}^2$ along $\mathbb{R}N$ and $(g - d\mathfrak{r}^2)$ along $N^\perp$.

Finally, $\mathfrak{i} \in (0, \infty]$ will stand for the injectivity radius of $T_R(\gamma_0)$, that is, for the minimum of the distance from a point p to its cut locus as p varies in $T_R(\gamma_0)$ [do Carmo 1992, pp. 267–273]. For each $r \in (0, R]$, the injectivity radius of $T_r(\gamma_0)$ will thus be *at least equal* to $\mathfrak{i}$. If M is compact, $\mathfrak{i}$ is finite, but $\mathfrak{i} = \infty$ if M is the hyperbolic space, for instance.

1B. ***Fermi charts and related notions.*** Let $n = \dim M$. Given an orthonormal basis $\{e_1, \dots, e_n\}$ of $T_{\gamma_0(0)}M$ with $e_n = (d\gamma_0/ds)(0)$, let us assign to each $p \in T_R(\gamma_0)$ the n-tuple $x = (\tilde{x}, x^n) \in \bar{B}^{n-1}(0, R) \times [0, \ell_0]$, where $\bar{B}^{n-1}(0, R)$ denotes the closure of the ball of radius R in the Euclidean space $\mathbb{R}^{n-1}$, given by $x(p) = (x^1, \dots, x^{n-1}, x^n)$ if and only if $v_0^\perp(p) = \sum_{\alpha=1}^{n-1} x^\alpha e_\alpha$ and $z(p) = x^n$. The map $x : T_R(\gamma_0) \to \bar{B}^{n-1}(0, R) \times [0, \ell_0]$ so defined is called a *Fermi chart* along the embedded geodesic γ_0. (In 1922, while a PhD student at the Scuola Normale Superiore in Pisa, motivated by the study of the equivalence principle in general relativity, Enrico Fermi was the first to consider such local coordinates, which he used along timelike paths; see [Gray and Vanhecke 1982, p. 217] and references therein.)

We see from this construction that $y = (\tilde{y}, y^n)$ is another such chart if and only if $y^n = x^n$ and there exists an orthogonal transformation $\mathfrak{R} \in \mathrm{O}(n-1)$ such that $\tilde{y} = \mathfrak{R}\tilde{x}$. The calculations which we will perform in the tube $T_R(\gamma_0)$ will be *invariant* (or tensorial) with respect to change of Fermi charts. We will freely use the local Euclidean metric $\mathfrak{e}_{\gamma_0} = \sum_{i=1}^n (dx^i)^2$ (just denoted by $\mathfrak{e}$, unless confusing) and the *affine structure* inherited from its (flat) Levi-Civita connection $D_{\gamma_0} = D$. The latter will be convenient to identify distinct tangent spaces and hence view vectors tangent to $T_R(\gamma_0)$ at distinct points as belonging to *the same* vector space. We will also view the Christoffel symbols $\Gamma_{ij}^k(x)$ of our original (global) connection ∇ as the components in the chart x of the local *tensorial* difference $(\nabla - D)$.

In the Fermi chart x, the components of the metric tensor g satisfy $g_{ij}(0, x^n) = \delta_{ij}$, $dg_{ij}(0, x^n) = 0$, so the Christoffel symbols vanish at $(0, x^n)$, meaning that g is osculating to $\mathfrak{e}$ along γ_0. We set $\|\cdot\|$ for the norm associated to the Euclidean metric $\mathfrak{e}$ and $\theta_0 = \min \|U\| \leqslant 1 \leqslant \Theta_0 = \max \|U\|$, where U runs over all unit[2] tangent vectors at points of $T_R(\gamma_0)$. For each $p \in T_R(\gamma_0)$, setting $\rho(x) = \sqrt{\sum_{\alpha=1}^{n-1} (x^\alpha)^2}$, we have $\mathfrak{r}(p) = \rho(x(p))$. The geodesic ray $t \in [0, 1] \to E_0(tv_0^\perp(p), z(p)) \in M$ reads $t \mapsto \mathcal{R}(t) = (tx^1, \dots, tx^{n-1}, x^n)$ with $x = x(p)$; being constant, its speed is equal to $\rho(x)$, so the unit vector field N reads $N(p) = \nu(x(p))$ with $\nu(x) = (1/\rho)(x) \sum_{\alpha=1}^{n-1} x^\alpha\, \partial/\partial x^\alpha$.

If $W = \sum_{i=1}^n W^i\, \partial/\partial x^i \in T_pM$, we may view W as a *constant* vector field in $T_R(\gamma_0)$, in other words, extend it to $T_R(\gamma_0)$ by D_{γ_0} parallelism, a notion well defined in any Fermi chart along γ_0. Following [Gray 2004, p. 21], let us call any such vector field a *Fermi field* (here, with respect to γ_0). Given a point $p \in T_R(\gamma_0)$ and vector field Z on $T_R(\gamma_0)$, we may similarly consider the Fermi field $Z(p)$, thinking of it as Z frozen at p. Among Fermi fields, one may distinguish those with $W^n = 0$ from those writing $Z = Z^n\, \partial/\partial x^n$ (sometimes called *axial*). For later use, we record the brackets identities

$$\left[\nu, \frac{\partial}{\partial x^n}\right] = 0 \quad \text{and} \quad \left[\nu, \rho\frac{\partial}{\partial x^\alpha}\right] = \frac{\partial\rho}{\partial x^\alpha}\nu \quad \text{for all } \alpha < n. \tag{2}$$

Finally, it will be convenient to consider on $T_R(\gamma_0)$ the field of projections $\Pi_0 = \sum_{\alpha=1}^{n-1} dx^\alpha \otimes \partial/\partial x^\alpha$, which is the constant (or Fermi) extension of the orthogonal projection of $T_{\gamma_0(0)}M$ onto $V_0^\perp$.

1C. *Estimates for geodesics in a thin tube.* Beforehand, let us recall a classical result, namely: there exists a continuous function $p \in M \to \chi(p) \in (0, \infty]$ called the *convexity radius*, which is smaller than the injectivity radius, such that, for each $\varrho \in (0, \chi(p))$, the Riemannian ball $B(p, \varrho)$ is strongly convex [Cheeger and Ebin 2008, pp. 103–105; Klingenberg 1995, pp. 84–85; Whitehead 1932]. For $r > 0$

[2] Here and below, to be understood for the metric g, unless otherwise specified.

small, we may thus consider the function $r \mapsto \chi_{\gamma_0}(r) = \min\{\chi(p),\, p \in T_r(\gamma_0)\}$, which is nonincreasing. We set $\mathfrak{c} = \chi_{\gamma_0}(R)$ and stress that $\mathfrak{c} \leqslant \mathfrak{i}$. Our first estimate is an upper bound on the length of the geodesics contained in the tube $T_{R_0}(\gamma_0)$ with $R_0 = \min(R, \mathfrak{c}/3)$.

Proposition 1.1. *If $\gamma : [0, \ell] \to \gamma(s) \in T_{R_0}(\gamma_0)$ is a unit-speed geodesic,[3] its length ℓ is bounded above by L_0, with $L_0 = \ell_0 + 2R$ if $\mathfrak{i} = \infty$, and $L_0 = 2(\ell_0 + \mathfrak{c})$ if $\mathfrak{c} < \infty$.*

Proof. If $\mathfrak{i} = \infty$, the geodesic γ is minimizing and unique in M. But we can join its endpoints $p = \gamma(0)$, $q = \gamma(\ell)$ by a geodesic path broken twice, namely, first by going along the geodesic ray from p to $\gamma_0(z(p))$, next by going from $\gamma_0(z(p))$ to $\gamma_0(z(q))$ along γ_0, then by going along the geodesic ray from $\gamma_0(z(q))$ to q. The total length of that broken path must be larger than ℓ and it is, indeed, at most equal to $L_0 = \ell_0 + 2R$.

If $\mathfrak{c} < \infty$, for each $\epsilon > 0$ small enough, the triangle inequality satisfied by the Riemannian distance on M shows that we can cover the tube $T_{R_0}(\gamma_0)$ by N open balls of radius $r = \mathfrak{c} - \epsilon$, successively centered at the points $\gamma_0(0), \gamma_0(r), \gamma_0(2r), \ldots, \gamma_0((N-1)r), \gamma_0(\ell_0)$, with $N = [\ell_0/\mathfrak{c}] + 1$. Now, the length of the restriction of the geodesic γ to each ball is bounded above by $2r$ and, letting $\epsilon \downarrow 0$, we obtain $\ell \leqslant 2N\mathfrak{c}$. □

Using a Fermi chart along γ_0, setting $R_1 = \frac{9}{10} R_0$, we can readily find a positive constant c_1 such that, for each $p \in T_{R_1}(\gamma_0)$, the following estimates hold at $x = x(p)$:

$$\|g - \mathfrak{e}\| \leqslant c_1 \rho^2(x), \qquad \|\nabla - D\| \leqslant c_1 \rho(x). \tag{3}$$

The purpose of our next proposition is twofold. On the one hand, it provides a radius under which the geodesics contained in a tube about γ_0 and longer than a given length $\delta > 0$ keep moving axially in a single direction; in particular, they must be embedded, like γ_0. On the other hand, it provides an estimate describing how C^0-close to γ_0 a geodesic should be in order to get C^1-close to it.

Proposition 1.2. *Fixing $\delta \in (0, L_0)$, let $r_1 > 0$ be given by*

$$r_1^2 \left(c_1 \Theta_0^2 + \frac{1}{\theta_0^2} \left(\frac{4}{\delta} + c_1 L_0 \Theta_0^2 \right)^2 \right) = 1.$$

For each $r \in (0, \min(R_1, r_1))$ and each unit-speed geodesic $s \in [0, \ell] \to \gamma(s) \in T_r(\gamma_0)$ with length $\ell \geqslant \delta$, the axial component $d\gamma^n/ds$ of the velocity cannot vanish. Moreover, the following estimate holds:

$$\left\| \varepsilon \frac{d\gamma}{ds} - \frac{\partial}{\partial x^n} \right\| \leqslant \left(\frac{4}{\ell} + c_1 \ell \Theta_0^2 \right) \rho_\gamma + \left(c_1 \Theta_0^2 + \frac{1}{\theta_0^2} \left(\frac{4}{\ell} + c_1 \ell \Theta_0^2 \right)^2 \right) \rho_\gamma^2, \tag{4}$$

[3] Throughout the paper, ℓ denotes the length of γ which may vary; it should be written $\ell(\gamma)$, of course, but we will stick to the short notation ℓ instead.

where ρ_γ *stands for* $\max_{\sigma\in[0,\ell]}\rho(\gamma(\sigma))$ *and* $\varepsilon = \pm 1$, *according to the sign of* $d\gamma^n/ds$.

Proof. Before proving the first assertion we require an estimate; namely, letting $s\in[0,\ell]\to\gamma(s)\in T_{R_1}(\gamma_0)$ be a unit-speed geodesic, we have

$$\left\|\Pi_0\frac{d\gamma}{ds}(s)\right\| \leqslant \left(\frac{4}{\ell}+c_1\ell\Theta_0^2\right)\rho_\gamma \quad \text{for all } s\in[0,\ell]. \tag{5}$$

Indeed, if $s\in\left[0,\frac{\ell}{2}\right]$, we write, for all $\alpha\in\{1,\dots,n-1\}$,

$$(\ell-s)\frac{d\gamma^\alpha}{ds}(s)=\gamma^\alpha(\ell)-\gamma^\alpha(s)-\int_s^\ell\int_s^S\frac{d^2\gamma^\alpha}{d\sigma^2}(\sigma)\,d\sigma\,dS,$$

while if $s\in\left[\frac{\ell}{2},\ell\right]$, we write instead

$$s\frac{d\gamma^\alpha}{ds}(s)=\gamma^\alpha(s)-\gamma^\alpha(0)-\int_0^s\int_s^S\frac{d^2\gamma^\alpha}{d\sigma^2}(\sigma)\,d\sigma\,dS.$$

In either case, transforming the last term of the right-hand side by means of the geodesic equation, recalling (3) and using the triangle and Schwarz inequalities, we readily infer (5). Writing

$$\left|\frac{d\gamma^n}{ds}\right|=\left\|\frac{d\gamma}{ds}\right\|\sqrt{1-\frac{\|\Pi_0\,d\gamma/ds\|^2}{\|d\gamma/ds\|^2}} \quad\text{and}\quad \left\|\frac{d\gamma}{ds}\right\|=\sqrt{1-(g-\mathfrak{e})\left(\frac{d\gamma}{ds},\frac{d\gamma}{ds}\right)},$$

the latter to be combined with (3), we get

$$\left|\frac{d\gamma^n}{ds}\right|\geqslant 1-c_1\rho_\gamma^2\Theta_0^2-\frac{1}{\theta_0^2}\left\|\Pi_0\frac{d\gamma}{ds}\right\|^2;$$

hence, using (5), we obtain the important lower bound

$$\left|\frac{d\gamma^n}{ds}(s)\right|\geqslant 1-\left(c_1\Theta_0^2+\frac{1}{\theta_0^2}\left(\frac{4}{\ell}+c_1\ell\Theta_0^2\right)^2\right)\rho_\gamma^2 \quad\text{for all } s\in[0,\ell]. \tag{6}$$

Recalling Proposition 1.1 and the assumption $\ell\geqslant\delta$, this shows that $d\gamma^n/ds$ cannot vanish provided the radius r of the tube in which the geodesic ranges satisfies

$$r^2\left(c_1\Theta_0^2+\frac{1}{\theta_0^2}\left(\frac{4}{\delta}+c_1L_0\Theta_0^2\right)^2\right)<1,$$

or else $r\in(0,\min(R_1,r_1))$, as we assumed. The first part of Proposition 1.2 is thus proved.

Moreover, letting now ε stand for the *sign* of $d\gamma^n/ds$, we have

$$|d\gamma^n/ds(s)|\equiv\varepsilon\,d\gamma^n/ds,$$

so we readily get from (6) and the obvious inequality $|d\gamma^n/ds| \leqslant \|d\gamma/ds\|$, the pinching

$$-\frac{1}{2}c_1\Theta_0^2\rho_\gamma^2 \leqslant 1-\varepsilon\frac{d\gamma^n}{ds} \leqslant \left(c_1\Theta_0^2+\frac{1}{\theta_0^2}\left(\frac{4}{\ell}+c_1\ell\Theta_0^2\right)^2\right)\rho_\gamma^2.$$

Combined with (5) this yields (4), since

$$\left\|\varepsilon\frac{d\gamma}{ds}(s)-\frac{\partial}{\partial x^n}\right\| \leqslant \left\|\Pi_0\frac{d\gamma}{ds}(s)\right\|+\left|\varepsilon\frac{d\gamma^n}{ds}(s)-1\right|. \qquad \square$$

Writing UM for the unit tangent bundle and $\mathrm{End}_s(TM)$ for the bundle of symmetric[4] endomorphisms of TM, let us consider the map

$$(p,U)\in UM \to \mathbb{J}(p,U)=R_p(\,\cdot\,,U)U \in \mathrm{End}_s(TM),$$

where R_p stands for the Riemann curvature tensor at the point $p\in M$. It satisfies $g(V,\mathbb{J}(p,U)W)\equiv S_p(V,U,W,U)$ where S_p stands for the sectional (or covariant Riemann) curvature tensor of the metric g at the point p; it is thus, indeed, *symmetric*. We denote by $\kappa^1(p,U)\leqslant\cdots\leqslant\kappa^{n-1}(p,U)$ the eigenvalues (each repeated with its multiplicity) of the nontrivial part of $\mathbb{J}(p,U)$, namely of its restriction to $U^\perp$. For each $\alpha\in\{1,\dots,n-1\}$, the map $(p,U)\in UM\to\kappa^\alpha(p,U)\in\mathbb{R}$ is C^1_{loc} [Kato 1995, pp. 122–123], hence uniformly Lipschitz for $p\in T_{R_0}(\gamma_0)$. So there exists a constant k_0 such that, for each pair $((p,U),(p',U'))\in UM^2$ with $\max(\mathfrak{r}_{\gamma_0}(p),\mathfrak{r}_{\gamma_0}(p'))\leqslant R_0$ and each $\alpha\in\{1,\dots,n-1\}$, the following uniform estimate holds:

$$\left|\kappa^\alpha(p,U)-\kappa^\alpha(p',U')\right| \leqslant k_0\big(d(p,p')+\|U-U'\|\big). \tag{7}$$

For each unit-speed geodesic $\sigma\in[0,\ell]\to\gamma(\sigma)\in M$, we write $s\mapsto\mathfrak{J}_\gamma(s)$ for the pullback to $[0,\ell]$ of the map $\mathbb{J}$ by the section

$$t\mapsto\left(\gamma(s),\frac{d\gamma}{d\sigma}(s)\right)\in UM$$

and call $\mathfrak{J}_\gamma(s)$ the *Jacobi operator* along the geodesic γ at s. We further write $\kappa_\gamma^1(s)\leqslant\cdots\leqslant\kappa_\gamma^{n-1}(s)$ for the eigenvalues of the restriction of $\mathfrak{J}_\gamma(s)$ to $\frac{d\gamma}{d\sigma}(s)^\perp$ and call them the *Jacobi curvatures* along γ at s.

Corollary 1.3. *Given δ and r as in Proposition 1.2, set*

$$k=k_0\left(1+\frac{4}{\delta}+c_1L_0\Theta_0^2+\left(c_1\Theta_0^2+\frac{1}{\theta_0^2}\left(\frac{4}{\delta}+c_1L_0\Theta_0^2\right)^2\right)r\right).$$

For each unit-speed geodesic $\sigma\in[0,\ell]\to\gamma(\sigma)\in T_r(\gamma_0)$ with length $\ell\geqslant\delta$ and

[4]Here, “unit” and “symmetric” refer to the Riemannian metric g, of course.

each $s \in [0, \ell]$, *the following estimate holds*:

$$\left|\kappa_\gamma^\alpha(s) - \kappa_0^\alpha(\gamma^n(s))\right| \leqslant k\rho_\gamma \quad \textit{for all } \alpha \in \{1, \ldots, n-1\},$$

where $\kappa_0^1 \leqslant \cdots \leqslant \kappa_0^{n-1}$ *stand for the Jacobi curvatures along* γ_0.

Proof. Fixing γ as stated, we may apply Proposition 1.2 to it. This yields an estimate for $\|(d\gamma/d\sigma)(s) - \partial/\partial x^n\|$ which, combined with the estimate (7) read at $(p, U) = \big(\gamma(s), (d\gamma/d\sigma)(s)\big)$ and $(p', U') = \big(\gamma_0(\gamma^n(s)), \partial/\partial x^n\big)$, yields the desired result. □

Corollary 1.3 shows in particular that, if the Jacobi operator along γ_0 stays *definite*, it must stay so (with the same signature) along geodesics longer than a given length and contained in a tube about γ_0 of small enough radius.

1D. *Family of Fermi maps near γ_0*. For each unit-speed geodesic $s \in [0, \ell] \to \gamma(s) \in T_{R_1}(\gamma_0)$, let $I_{\gamma_0}(\gamma) \subset [0, \ell_0]$ denote the axial image interval $\gamma^n([0, \ell])$ and $T(\gamma_0, \gamma)$, the shortest piece of tube about γ_0 containing γ, equal to $\{m \in T_{\rho_\gamma}(\gamma_0),\ x^n(m) \in I_{\gamma_0}(\gamma)\}$. If such a geodesic γ is an embedding, when is it possible to construct a Fermi map along it such that a point $m \in T(\gamma_0, \gamma)$ may stay *outside* the corresponding tube about γ if and only if its height $z_\gamma(m)$ relative to γ satisfies either $z_\gamma(m) < 0$ or $z_\gamma(m) > \ell$? When such a possibility occurs, we call (γ_0, γ)-*exceptional* the latter points and (γ_0, γ)-*accessible* all other points of $T(\gamma_0, \gamma)$. Sticking to the notations of Proposition 1.2, we will prove the following:

Proposition 1.4. *For each* $\delta \in (0, \ell_0)$, *there exists* $r_2 \in (0, \min(R_1, r_1))$ *such that*, *for each unit-speed geodesic* γ *longer than* δ *and contained in* $T_{r_2}(\gamma_0)$, *a Fermi map can be constructed along* γ *with corresponding tube about* γ *containing the whole of* $T(\gamma_0, \gamma)$ *but its* (γ_0, γ)*-exceptional points.*

We call *family of Fermi maps near* γ_0 the map which assigns, to each unit-speed geodesic γ as stated and each (γ_0, γ)-accessible point $m \in T_{r_2}(\gamma_0)$, the image of m by the Fermi map along γ.

Proof. The idea is to use a suitable implicit function theorem argument along γ_0. Since it is absent from the literature, we will present it carefully. Let us fix $\delta \in (0, \ell_0)$ and a unit-speed geodesic $\sigma \in [0, \ell^*] \to \gamma^*(\sigma) \in T_{r_2}(\gamma_0)$, with $\ell^* \geqslant \delta$ and $r_2 \in (0, \min(R_1, r_1))$ to be chosen later. From Proposition 1.2, we know that γ^* is an embedding. We can thus construct a tube $T_\varrho(\gamma^*)$ about γ^*, for some radius $\varrho > 0$, as done for γ_0 in Section 1A. We want $\rho_{\gamma^*} \leqslant r_2$ small enough compared to ϱ such that the tube $T_\varrho(\gamma^*)$ contains $T(\gamma_0, \gamma^*)$ but its exceptional points. Can we choose the radius r_2 such that this property holds for every such geodesic γ^*?

First, we observe that the required property holds for γ^* if and only if it holds for the reversed geodesic γ^*_{rev}, given by $\sigma \in [0, \ell^*] \to \gamma^*_{\text{rev}}(\sigma) = \gamma^*(\ell^* - \sigma)$. Therefore, applying Proposition 1.2 to γ^*, we may assume with no loss of generality that $d\gamma^{*n}/d\sigma$ is positive.

Next, we note that the geodesic γ^* is given by its Cauchy data

$$(p^*, u^*) = \left(\gamma^*(0), \frac{d\gamma^*}{d\sigma}(0)\right) \in UM$$

and its length $\ell^* \in [\delta, L_0]$, while the generic point m^* of the tube $T_\varrho(\gamma^*)$ is determined by its Fermi map image $F_{\gamma^*}(m^*)$, namely by its height $\sigma^* = z_{\gamma^*}(m^*) \in [0, \ell^*]$ and by the vector $V^* = v^\perp_{\gamma^*}(m^*) \in (u^*)^\perp$ such that $|V^*| \leqslant \varrho$ and $E_{\gamma^*}(V^*, \sigma^*) = m^*$. Here, we have denoted by $E_{\gamma^*} : (u^*)^\perp \times (-\epsilon, \ell^* + \epsilon) \to M$ (respectively, by $v^\perp_{\gamma^*}$) the analogue for γ^* of the map E_0 (respectively, of the component $v^\perp_0$) defined for γ_0 at the beginning of Section 1A.

The resulting point (p^*, u^*, V^*), the amalgam of the Cauchy data of γ^* with the Fermi component $V^* = v^\perp_{\gamma^*}(m^*) \in (u^*)^\perp$ of m^*, lies in the vector bundle $\ker T\pi \to UM$, the kernel of the tangent map to the natural projection $\pi : UM \to M$. Sticking to the Fermi chart x along γ_0, we use it to build a chart of $\ker T\pi$ near (p^*, u^*, V^*) by assigning to each neighboring point (p, u, V) the $(3n-2)$-tuple $(x^1, \dots, x^n, u^1_0, \dots, u^{n-1}_0, V^1_0, \dots, V^{n-1}_0)$ with $x^i = x^i(p)$ and u^α_0, V^α_0 defined as follows. Firstly, for each tangent vector $W \in T_pM$, let $\overline{W}_0 \in T_{p^\perp_0}M$, with $p^\perp_0 = p^\perp_{\gamma_0} \equiv \gamma_0(x^n(p))$, denote its (backward) parallel transport[5] along the geodesic ray $[\gamma_0, p]$, and $W_0 \in T_{\gamma_0(0)}M$, similarly from the latter now along γ_0. We pause to record a lemma (the proof of which is left as an easy exercise):

Lemma 1.5. *If U is a unit tangent vector at $p \in T_{R_1}(\gamma_0)$ and $\overline{U}_0$ stands for its parallel transport to the point $\gamma_0(x^n(p))$ along the geodesic ray $[\gamma_0, p]$, the following estimate holds*:

$$\|U - \overline{U}_0\| \leqslant c_1 \Theta_0 \mathfrak{r}^2(p).$$

Applying this lemma, combined with Proposition 1.2 and the triangle inequality, to the vector $u^* \in T_{p^*}M$, and recalling that $\|\cdot\| \equiv |\cdot|$ along γ_0, we infer the estimate

$$|u^*_0 - e_n| \leqslant k_1 r_2, \tag{8}$$

with

$$k_1 = \frac{4}{\delta} + c_1 L_0 \Theta_0^2 + \left(c_1\Theta_0 + c_1\Theta_0^2 + \frac{1}{\theta_0^2}\left(\frac{4}{\delta} + c_1 L_0 \Theta_0^2\right)^2\right) r_1.$$

Here, we used the positivity assumption made above on $(u^*)^n$. Taking $r_2 < 1/k_1$, this estimate implies the positivity of $(u^*_0)^n$. Back to the definition of the chart of

[5] Henceforth, with respect to the Levi-Civita connection ∇, unless otherwise specified.

$\ker T\pi$ under elaboration, we take (p, u, V) close enough to (p^*, u^*, V^*) for u_0^n to be still positive, and we define the u_0^α's and V_0^α's by

$$\sum_{\alpha=1}^{n-1} u_0^\alpha e_\alpha = \Pi_0 u_0, \quad \sum_{\alpha=1}^{n-1} V_0^\alpha e_\alpha = \Pi_0 V_0.$$

We recover the full parallel transported vectors u_0, V_0, by setting

$$u_0^n = \sqrt{1 - \sum_{\alpha=1}^{n-1} (u_0^\alpha)^2},$$

since $|u_0| = 1$ and $u_0^n > 0$, and

$$V_0^n = -\frac{1}{u_0^n} \sum_{\alpha=1}^{n-1} u_0^\alpha V_0^\alpha,$$

since $V_0 \perp u_0$. So $(x^i, u_0^\alpha, V_0^\alpha)$ is, indeed, a local chart of $\ker T\pi$. Although heavier, let us denote it rather by $(x^{*i}, u_0^{*\alpha}, V_0^{*\alpha})$ since we are now willing to move around the geodesic γ^* and the point $m^* \in T_\varrho(\gamma^*)$, hence to let the point (p^*, u^*, V^*) itself vary in $\ker T\pi$ near $(p_0, u_0, V_0) = \big(\gamma_0(s_0), (d\gamma_0/ds)(s_0), 0\big)$ with $s_0 \in [0, \ell_0]$. Deferring the completion of the present proof, we pause to set up an appropriate implicit function theorem.

Implicit function theorem argument. In this section, the requirement that the geodesic γ^* be longer than δ will be unnecessary, thus ignored provisionally. Given $s_0 \in [0, \ell_0]$ and $\sigma_0 \in [0, \ell_0 - s_0]$, let the point $(p^*, u^*, V^*) \in \ker T\pi$ be close to (p_0, u_0, V_0) and the real $\sigma^* \in \mathbb{R}^+$ be close to σ_0; let a further point m belong to $T_{r_2}(\gamma_0)$. Setting $\gamma^*(\sigma) = \exp_{p^*}(\sigma u^*)$ and $m^* = E_{\gamma^*}(V^*, \sigma^*)$, consider the map

$$\Psi(p^*, u^*, V^*, \sigma^*, m) = x(m^*) - x(m) \in \mathbb{R}^n.$$

Using the chart $(x^{*i}, u_0^{*\alpha}, V_0^{*\alpha})$ for (p^*, u^*, V^*) and the chart x^i for m, let us denote the local expression of Ψ (respectively, $x \circ E_{\gamma^*}$) by

$$\Psi^i(x^{*j}, u_0^{*\alpha}, V_0^{*\alpha}, \sigma^*, x^j) = E^i(x^{*j}, u_0^{*\alpha}, V_0^{*\alpha}, \sigma^*) - x^i.$$

At the point given by[6] $x^{*\alpha} = 0$, $x^{*n} = s_0$, $u_0^{*\alpha} = 0$, $V_0^{*\alpha} = 0$, $\sigma^* = \sigma_0$, $x^\alpha = 0$, $x^n = s_0 + \sigma_0$, we have

$$\Psi^i\big((\vec{0}, s_0), \vec{0}, \vec{0}, \sigma_0, (\vec{0}, s_0 + \sigma_0)\big) = 0 \quad \text{for all } i \in \{1, \dots, n\}$$

and

$$\det\left(\frac{\partial \Psi^j}{\partial(V_0^{*\alpha}, \sigma^*)}\big((\vec{0}, s_0), \vec{0}, \vec{0}, \sigma_0, (\vec{0}, s_0 + \sigma_0)\big)\right) \neq 0,$$

[6] Throughout with α ranging in $\{1, \dots, n-1\}$.

where $\vec{0}$ stands for the zero vector of $\mathbb{R}^{n-1}$. The latter equation holds since

$$\frac{\partial \Psi^j}{\partial (V_0^{*\alpha}, \sigma^*)} \equiv \frac{\partial E^j}{\partial (V_0^{*\alpha}, \sigma^*)}$$

and $dE^j((\vec{0}, s_0), \vec{0}, \vec{0}, \sigma_0) \equiv dx^j \circ dE_0(0, s_0 + \sigma_0)$, where $dE_0(0, s_0 + \sigma_0)$ is an *isomorphism* as seen in Section 1A. We are thus in position to apply the implicit function theorem [Lang 2002]. There exists a real $\epsilon > 0$ and a unique map $(x^{*j}, u_0^{*\alpha}, x^j) \to F^* = (\mathcal{V}_0^{*1}, \ldots, \mathcal{V}_0^{*n-1}, \varsigma^*)$ such that, if

(9) $\rho(x^*) \leqslant \epsilon, \quad |x^{*n} - s_0| \leqslant \epsilon, \quad |\Pi_0 u_0^*| \leqslant \epsilon, \quad \rho(x) \leqslant \epsilon, \quad |x^n - (s_0 + \sigma_0)| \leqslant \epsilon,$

the identities

$$\Psi^i\big(x^{*j}, u_0^{*\alpha}, \mathcal{V}_0^{*\alpha}(x^{*k}, u_0^{*\alpha}, x^k), \varsigma^*(x^{*k}, u_0^{*\alpha}, x^k), x^j\big) \equiv 0 \quad \text{for all } i \in \{1, \ldots, n\}$$

are satisfied with $\sum_{\alpha=1}^{n-1} (\mathcal{V}_0^{*\alpha}(x^{*k}, u_0^{*\alpha}, x^k))^2$ and $\big|\varsigma^*(x^{*k}, u_0^{*\alpha}, x^k) - \sigma_0\big|$ small. By construction, these identities imply $m = m^*$; in other words, the map

$$x^j \to F^{*i}(x^{*j}, u_0^{*\alpha}, x^j)$$

is nothing but the expression of the Fermi map F_{γ^*} along the geodesic $\gamma^*(\sigma) = \exp_{p^*}(\sigma u^*)$ read in the Fermi chart x along γ_0. Finally, let us stress that the real $\epsilon > 0$ occurring in (9) may be chosen so small that it becomes independent of the pair of parameters (s_0, σ_0), because the latter lies in a compact subset of $\mathbb{R}^2$, namely in the triangle of the positive quadrant given by $s_0 + \sigma_0 \leqslant \ell_0$. Henceforth, we fix $\epsilon > 0$ so.

Completion of the proof of Proposition 1.4. Back to the case of our previous geodesic γ^*, supposed longer than δ and with positive axial component, we are now in position to choose the radius r_2 of the tube about γ_0 in which γ^* should lie. First of all, we fix a point $m \in T(\gamma_0, \gamma^*)$. So far, we have required $r_2 \in (0, \min(R_1, r_1, 1/k_1))$. Redoing the preceding implicit function theorem argument now with $p^* = \gamma^*(0)$, $s_0 = x^n(p^*)$, $s_0 + \sigma_0 = x^n(m)$, the first and fourth inequalities of (9) prompt us to take $r_2 \leqslant \epsilon$. Besides, we must further shrink $r_2 > 0$ in order to keep γ^* nearly vertical so that the third inequality of (9) holds as well. From (8), we can do it by taking $r_2 \leqslant \epsilon/k_1$, as easily verified. Altogether, if the geodesic γ^* is longer than $\delta \in (0, \ell_0)$ with $d\gamma^{*n}/d\sigma > 0$ and if it is contained in the tube $T_{r_2}(\gamma_0)$ with $r_2 \in (0, \min(R_1, r_1, \epsilon/k_1))$, the triple

$$\left(x^{*i} = x^{*i}(\gamma^*(0)),\ u_0^{*\alpha} = u_0^{*\alpha}\left(\frac{d\gamma^*}{d\sigma}(0)\right),\ x^i = x^i(m)\right)$$

satisfies the bounds (9). So we may consider its image by the local map F^* precedingly constructed. In particular, it follows that the point m lies in a tube about the embedded geodesic γ^* if and only if its height $z_{\gamma^*}(m) = \varsigma^*(x^{*i}, u_0^{*\alpha}, x^i)$

lies in the interval $[0, \ell^*]$. Since the point m was arbitrarily fixed in $T(\gamma_0, \gamma^*)$, we are done. $\square$

1E. *Second fundamental form of a cylinder.* If $n > 2$, sticking to the notations of Section 1A, let us study the second fundamental form of a cylinder $C_r(\gamma_0)$ of small radius r about γ_0.

Proposition 1.6. *Given $r \in (0, \min(1, R))$, a point $p \in C_r(\gamma_0)$ and a pair of vectors $(V, W) \in T_pC_r(\gamma_0) \times T_pC_r(\gamma_0)$, let us denote by $\mathrm{II}_p(V, W)$ the second fundamental form of the cylinder $C_r(\gamma_0)$ calculated at p on (V, W). If we extend the vectors V, W and $N(p)$ as Fermi fields on $T_R(\gamma_0)$ and set $p^\perp = \gamma_0(z(p))$, the following asymptotic expansion holds*:

$$\begin{aligned}\mathrm{II}_p(V, W) = -\frac{1}{r} g(\Pi_0 V, \Pi_0 W)(p^\perp) + r\big(&S(V, N(p), W, N(p))(p^\perp)\\ &- \tfrac{1}{3} S(\Pi_0 V, N(p), \Pi_0 W, N(p))(p^\perp)\big) + O(r^2),\end{aligned}$$

where, again, S stands for the sectional curvature tensor.

Proof. By definition [Gray 2004, p. 33; do Carmo 1992, p. 128], we have $\mathrm{II}_p(V, W) = g(-\nabla_V N, W)(p)$ and, here, one may allow the vectors V, W to be arbitrary in T_pM since N is a vector field defined outside $C_r(\gamma_0)$. Covariant differentiation of the generalized Gauss lemma identity $g(N, \cdot) = d\mathfrak{r}$ on $\{\mathfrak{r} > 0\} \subset T_R(\gamma_0)$ yields

$$\mathrm{II}_p(V, W) = -\nabla d\mathfrak{r}(V, W)(p). \tag{10}$$

More generally, for each pair of vector fields (A, B), we find $\nabla d\mathfrak{r}(A, B) = g(A, \nabla_B N) = g(B, \nabla_A N)$; hence also, using Lie brackets,

$$2\nabla d\mathfrak{r}(A, B) = N.g(A, B) + g(A, [B, N]) + g(B, [A, N]), \tag{11}$$

since ∇ is torsionless. Taking a Fermi chart x along γ_0 such that

$$x(p) = (r, \underbrace{0, \ldots, 0}_{n-2}, x^n(p)),$$

let us calculate $\nabla d\rho(r, 0, x^n)$ using (11) with A and B equal to the $\partial/\partial x^i$. Note that $\nu(r, 0, x^n) = \partial/\partial x^1$ and $d\rho(r, 0, x^n) = dx^1$. From (1), we get $g_{1i}(r, 0, x^n) = \delta_{1i}$ and $N \cdot g(\partial/\partial x^1, \partial/\partial x^i)(r, 0, x^n) = 0$. From (2), we find $[\partial/\partial x^n, \nu](r, 0, x^n) = 0$ and

$$\left[\frac{\partial}{\partial x^\alpha}, \nu\right](r, 0, x^n) = \frac{1}{r}\left(\frac{\partial}{\partial x^\alpha} - \delta_{1\alpha}\frac{\partial}{\partial x^1}\right) \quad \text{for all } \alpha < n;$$

in particular, $[\partial/\partial x^1, \nu](r, 0, x^n) = 0$. Besides, for $i, j \in \{2, \ldots, n\}$, we can derive the local expressions of $N.g(\partial/\partial x^i, \partial/\partial x^j)(r, 0, x^n) = (\partial g_{ij}/\partial x^1)(r, 0, x^n)$ from the following Riemann-type formulas extended to the Fermi setting [Spivak 1979; Delanoë and Ge 2010, Lemma 2]:

$$g_{ab}(x^1, \underbrace{0, \dots, 0}_{n-2}, x^n) = \delta_{ab} - \tfrac{1}{3}(x^1)^2 R_{a1b1}(\underbrace{0, \dots, 0}_{n-1}, x^n) + O((x^1)^3),$$

with $a, b \in \{2, \dots, n-1\}$, and

$$\begin{aligned} g_{an}(x^1, 0, x^n) &= -\tfrac{2}{3}(x^1)^2 R_{a1n1}(0, x^n) + O((x^1)^3), \\ g_{nn}(x^1, 0, x^n) &= 1 - (x^1)^2 R_{n1n1}(0, x^n) + O((x^1)^3), \end{aligned} \tag{12}$$

where x^1 stands for a small real parameter and R_{ijkl} for the components of the sectional curvature tensor. Doing so, we obtain the expression

$$\begin{aligned} \nabla d\rho(r, 0, x^n) = \sum_{a=2}^{n-1}\sum_{b=2}^{n-1}\Big(\frac{1}{r}\delta_{ab} - \tfrac{2}{3} r R_{a1b1}(0, x^n) + O(r^2)\Big)\, dx^a \otimes dx^b \\ + \sum_{a=2}^{n-1}\big(-r R_{a1n1}(0, x^n) + O(r^2)\big)(dx^a \otimes dx^n + dx^n \otimes dx^a) \\ + \big(-r R_{n1n1}(0, x^n) + O(r^2)\big)\, dx^n \otimes dx^n. \end{aligned} \tag{13}$$

The latter combined with (10) yields the proposition. ☐

Remark 1.7. For later use, we record here that, if $n = 2$, recalling (1), the expansion of the metric in the Fermi chart x becomes simply

$$g(x^1, x^2) = dx^1 \otimes dx^1 + \big(1 - (x^1)^2 K(0, x^2) + O((x^1)^3)\big)\, dx^2 \otimes dx^2,$$

where K stands for the Gauss curvature of M. Accordingly, still from (11), the Hessian formula (13) becomes

$$\nabla d\rho(r, x^2) = \big(-r K(0, x^2) + O(r^2)\big)\, dx^2 \otimes dx^2.$$

2. Further properties when the Jacobi operator is negative

From the properties established is the preceding section for a thin tube about the geodesic γ_0, we will now derive stronger ones by assuming that the operator $\mathcal{J}_{\gamma_0}$ is negative, as done in Theorem 0.2. Specifically, using the notations of Corollary 1.3 and setting $\bar\kappa_0 = \max_{s\in[0,\ell_0]} \kappa_0^{n-1}(s)$, our assumption means that $\bar\kappa_0 < 0$; henceforth, it is implicitly made.

Proposition 2.1 (the second fundamental form stays definite). *One can find a small real $r_3 > 0$ such that, for each $p \in T_{r_3}(\gamma_0)$ with $r = \mathfrak{r}(p) \neq 0$, the second fundamental form of $C_r(\gamma_0)$ at the point p is negative definite.*

Proof. Let us take a Fermi chart x at the point p like the one used in the proof of Proposition 1.6 and write with it the expression of $\mathrm{II}_p(V, W)$ found in that proposition, with $V = W = \sum_{i=2}^{n} V^i \partial/\partial x^i \in T_p C_r(\gamma_0)$. We find

$$
\begin{aligned}
(14)\quad & \mathrm{II}_p(V, V) \\
&= -\frac{1}{2r}\sum_{a=2}^{n-1}(V^a)^2 + \frac{r}{2}R_{n1n1}(0, x^n)(V^n)^2 \\
&\quad - \frac{1}{4r}\left(\sum_{a=2}^{n-1}(V^a)^2 - 8r^2\sum_{a=2}^{n-1}R_{a1n1}(0, x^n)V^aV^n - 2r^2R_{n1n1}(0, x^n)(V^n)^2\right) \\
&\quad - \frac{1}{4r}\sum_{a=2}^{n-1}\sum_{b=2}^{n-1}V^aV^b\big(\delta_{ab} - \tfrac{8}{3}r^2R_{a1b1}(0, x^n)\big) + O(r^2)
\end{aligned}
$$

and the result readily follows from $R_{n1n1}(0, x^n) \leqslant \bar{\kappa}_0 < 0$, provided r is taken small enough. □

Proposition 2.2 (geodesics obey a maximum principle). *One can find a small real $r_4 > 0$ such that, for each geodesic path $t \in [0, 1] \to \gamma(t) \in T_{r_4}(\gamma_0)$, the following inequality holds*:

$$\max_{t\in[0,1]} \mathfrak{r}(\gamma(t)) \leqslant \max\big(\mathfrak{r}(\gamma(0)), \mathfrak{r}(\gamma(1))\big).$$

Moreover, if $\mathfrak{r}(\gamma(\vartheta)) = \max\big(\mathfrak{r}(\gamma(0)), \mathfrak{r}(\gamma(1))\big)$ for some $\vartheta \in (0, 1)$, the path γ must be constant.

Proof. Anytime $t \in [0, 1] \to \gamma(t) \in T_R(\gamma_0)$ is a geodesic, at each $t \in [0, 1]$ such that $\mathfrak{r}(\gamma(t)) \neq 0$, we have

$$\frac{d^2}{dt^2}\big(\mathfrak{r}(\gamma(t))\big) = \nabla d\mathfrak{r}(\gamma(t))\left(\frac{d\gamma}{dt}, \frac{d\gamma}{dt}\right).$$

If $n > 2$, combining (13) with (14) written with $V = \frac{d\gamma}{dt}$, we infer that the second derivative of the auxiliary real function $t \in [0, 1] \to \mathfrak{r}(\gamma(t))$ is *nonnegative* on $[0, 1]$ provided $\mathfrak{r}(\gamma(t)) \leqslant r_4 = r_3$. If $n = 2$, the same conclusion holds with r_4 small enough, due to Remark 1.7 read with $K(0, x^2) \leqslant \bar{\kappa}_0 < 0$. In any case, the maximum principle [Protter and Weinberger 1967] implies the first part of the proposition. Moreover, it yields $\mathfrak{r} \circ \gamma \equiv \mathfrak{r}(\gamma(\vartheta)) =: r_\vartheta > 0$; hence $(d\gamma/dt)(t) \in T_{\gamma(t)}C_{r_\vartheta}(\gamma_0)$ for each $t \in [0, 1]$. From (10) and Proposition 2.1 combined with

$$\frac{d^2}{dt^2}\big(\mathfrak{r}(\gamma(t))\big) \leqslant 0,$$

we infer that $d\gamma/dt \equiv 0$, so γ must indeed be constant. □

Before moving on to the next property, we state a lemma of independent interest. (A reader parachuting to this point should understand it preceded by: "*Let γ_0 be an embedded unit-speed geodesic with negative Jacobi operator.*")

Lemma 2.3. *One can find a small real $r_5 > 0$ such that the inequality $g \geqslant d\mathfrak{r}_{\gamma_0}^2 + dz_{\gamma_0}^2$ between quadratic forms holds at each point of $\{p \in T_{r_5}(\gamma_0), \mathfrak{r}(p) > 0\}$.*

Proof. Take a point p as stated and a Fermi chart x along γ_0 such that $x(p) = (r, 0, \dots, 0, x^n)$. From Remark 1.7 read with $K(0, x^2) \leqslant \bar{\kappa}_0 < 0$, the lemma appears straightforward if $n = 2$. In higher dimensions, from (1) and the expansion of $g_{ij}(x^1, 0, x^n)$ in (12), we infer that, for each vector $V = \sum_{i=1}^n V^i\, \partial/\partial x^i \in T_pM$, the quadratic form $(g - d\mathfrak{r}_{\gamma_0}^2 - dz_{\gamma_0}^2)(p)$ applied to V can be expressed in the chart x, up to $O(r^3)$ terms, as the sum of two quadratic polynomials in V, namely $\sum_{a,b=2}^{n-1}\bigl(\frac{1}{2}\delta_{ab} - \frac{1}{3}r^2 R_{a1b1}(0, x^n)\bigr)V^aV^b$ and

$$\sum_{a=2}^{n-1}\bigl(\tfrac{1}{2}V_a^2 - \tfrac{4}{3}r^2 R_{a1n1}(0, x^n)V^aV^n\bigr) - r^2 R_{n1n1}(0, x^n)(V^n)^2.$$

By taking $r > 0$ small enough, and using $R_{n1n1}(0, x^n) \leqslant \bar{\kappa}_0 < 0$ for the second polynomial, we can make each polynomial nonnegative. □

Proposition 2.4 (γ_0 is minimizing). *Take $r_5 > 0$ as in Lemma 2.3. The length of each piecewise C^1 path $t \in [0, 1] \to c(t) \in M$ ranging in $T_{r_5}(\gamma_0)$ with $z(c(0)) = 0$ and $z(c(1)) = \ell_0$ must be at least equal to ℓ_0. Furthermore, if equality holds and $\mathfrak{r} \circ c(t) = 0$ for some $t \in [0, 1]$ then c, reparametrized by an arc-length parameter suitably shifted to avoid jumps*[7] *on each subinterval of* $[0, 1]$ *in the interior of which c is C^1 and $dc/dt \neq 0$, coincides with γ_0.*

Proof. Let c be a path as stated and x a Fermi chart along γ_0. From Lemma 2.3, the length of c satisfies

$$\ell \geqslant \int_0^1 \sqrt{\left(\frac{d}{dt}(\rho \circ c)\right)^2 + \left(\frac{dc^n}{dt}\right)^2}\, dt.$$

Therefore, if $\int_0^1 \bigl|(d/dt)(\rho \circ c)\bigr|\, dt \neq 0$, we have $\ell > \int_0^1 |dc^n/dt|\, dt \geqslant \ell_0$ as asserted. Moreover, if $\ell = \ell_0$, we see that $(d/dt)(\rho \circ c)$ must vanish, hence also $(\rho \circ c)$ anytime it does at some $t \in [0, 1]$. In that case, the images of c and γ_0 coincide, so $|dc/dt| = \|dc/dt\| = |dc^n/dt|$ and $\int_0^1 |dc^n/dt|\, dt = \ell_0 = c^n(1) - c^n(0) = \int_0^1 (dc^n/dt)\, dt$. The latter equality implies that $dc^n/dt \geqslant 0$, so the path c, reparametrized by arc length as stated, must indeed coincide with γ_0. □

[7] By taking the initial value of the parameter on a subinterval equal to (zero, of course, on the first subinterval and elsewhere to) the final value of the parameter on the preceding subinterval.

Proposition 2.5 (long geodesics have a negative Jacobi operator). *Given $\delta > 0$, we can find $r_6 \in (0, \min(R_1, r_1)]$ such that, for each $r \in (0, r_6)$ and each unit-speed geodesic $\sigma \in [0, \ell] \to \gamma(\sigma) \in T_r(\gamma_0)$ with length $\ell \geqslant \delta$, the Jacobi operator $\mathcal{J}_\gamma$ is negative, or else* $\max_{s\in[0,\ell]} \kappa_\gamma^{n-1}(s) < 0$.

Proof. Let $k = k(r)$ be the affine function of r defined in Corollary 1.3 and r^+ be the positive root of the quadratic equation $rk(r) + \bar{\kappa}_0 = 0$; the proposition holds with $r_6 = \min(R_1, r_1, r^+)$ by Corollary 1.3. □

Proposition 2.6 (each geodesic is minimizing). *One can find a small real $r_7 > 0$ such that, for each unit-speed geodesic $s \in [0, \ell] \to \gamma(s) \in M$ and each piecewise C^1 path $t \in [0, 1] \to c(t) \in M$, both ranging in $T_{r_7}(\gamma_0)$ with $c(0) = \gamma(0)$, $c(1) = \gamma(\ell)$, the length of c must be at least equal to ℓ. Moreover, equality holds if and only if c, reparametrized by a suitable arc length parameter on each subinterval of* $[0, 1]$ *in the interior of which c is C^1 and $dc/dt \neq 0$, coincides with γ.*

Proof. Let γ be a geodesic of length ℓ as stated. The proposition is obvious if $\ell < \mathfrak{i}$. If $\ell \geqslant \mathfrak{i}$, which we suppose in the proof, we may use Propositions 1.2 and 1.4 read with $\delta = \mathfrak{i}$; the radii r_1 and r_2 are understood accordingly and we take $r_7 \leqslant r_2$. In this situation, we know that γ is an embedding and there exists a Fermi chart x_γ along γ whose domain $T_\eta(\gamma)$ contains $T(\gamma_0, \gamma)$ but the (γ_0, γ)-exceptional points.

Our next task is the main one; namely, we must specify how the radius η of that tubular domain can be controlled by r_7. By inspecting the proof of Proposition 1.4, we see (sticking to its notations, except for γ^* now written γ, so $m^* = \gamma(0)$, $u^* = (d\gamma/ds)(0)$) that such a control amounts to a similar one on

$$\left\|\mathcal{V}_0^*(x^*, \Pi_0 u_0^*, x)\right\|^2 = \sum_{i=1}^{n} \left(\mathcal{V}_0^{*i}(x^*, \Pi_0 u_0^*, x)\right)^2,$$

where x^*, $\Pi_0 u_0^*$, x satisfy the bounds (9) now read with $\epsilon = r_7$ and where $\mathcal{V}_0^{*n}$ has to be defined by

$$\mathcal{V}_0^{*n} = -\frac{1}{u_0^{*n}} \sum_{\alpha=1}^{n-1} u_0^{*\alpha} \mathcal{V}_0^{*\alpha} \quad \text{with } u_0^{*n} = \pm\sqrt{1 - \sum_{\alpha=1}^{n-1} (u_0^{*\alpha})^2}.$$

Furthermore, as $r_7 \downarrow 0$, we know that $\sum_{\alpha=1}^{n-1} (\mathcal{V}_0^{*\alpha})^2$ tends to zero. All we require is thus a uniform positive lower bound on $|u_0^{*n}|$. Such a bound will follow from (6) and Lemma 1.5. Indeed, the former combined with Proposition 1.1 implies here that

$$\left|\frac{d\gamma^n}{ds}\right| \geqslant 1 - \left(c_1 \Theta_0^2 + \frac{1}{\theta_0^2}\left(\frac{4}{\mathfrak{i}} + 2c_1 \Theta_0^2(\ell_0 + \mathfrak{i})\right)^2\right) r_7^2,$$

which in turn yields $|u_0^{*n}| \geqslant |d\gamma^n/ds| - c_1 \Theta_0^2 r_7^2$. Thus we get

$$|u_0^{*n}| \geqslant 1 - \left(2c_1\Theta_0^2 + \frac{1}{\theta_0^2}\left(\frac{4}{\mathfrak{i}} + 2c_1\Theta_0^2(\ell_0 + \mathfrak{i})\right)^2\right) r_7^2.$$

Defining $\mathfrak{r}_1 > 0$ by, say,

$$\mathfrak{r}_1^2\left(2c_1\Theta_0^2 + \frac{1}{\theta_0^2}\left(\frac{4}{\mathfrak{i}} + 2c_1\Theta_0^2(\ell_0 + \mathfrak{i})\right)^2\right) = \frac{1}{2},$$

and taking $r_7 \leqslant \mathfrak{r}_1$, we obtain $|u_0^{*n}| \geqslant \frac{1}{2}$. Now, it is clear that $\|\mathcal{V}_0^*(x^*, \Pi_0 u_0^*, x)\|$ tends to zero as $r_7 \downarrow 0$. Here, among the arguments of $\mathcal{V}_0^*$, we are given the first one, since $x^* = x(\gamma(0))$; similarly for the second one, since $\Pi_0 u_0^*$ is defined out of $(d\gamma/ds)(0)$; the sole variable is the third one, since $x = x(m)$ with $m \in T(\gamma_0, \gamma) \cap T_\eta(\gamma)$. Moreover, using the aforementioned Fermi chart x_γ, the identity

$$\rho(x_\gamma) = \|\mathcal{V}_0^*(x^*, \Pi_0 u_0^*, x)\|$$

holds. So $\rho(x_\gamma) \downarrow 0$ as $r_7 \downarrow 0$, which shows that the implicit function theorem used in the proof of Proposition 1.4 allows us to let η go to zero as $r_7 \downarrow 0$.

Besides, Proposition 2.5 read with $\delta = \mathfrak{i}$ implies that, if we take $r_7 < r_6$, the Jacobi operator of γ is negative.

We conclude that there exists $r_7 > 0$ small enough such that, if γ ranges in $T_{r_7}(\gamma_0)$, the radius η of the tube about γ provided by Proposition 1.4 may be taken small enough such that Lemma 2.3 and Proposition 2.4 hold for the geodesic γ in $T_\eta(\gamma)$.

Now, we are in position to complete the proof of Proposition 2.6. Let c be a path as stated. By the definition of $T(\gamma_0, \gamma)$, the smallness of r_7 (hence of η) and the property of $T_\eta(\gamma)$ proved in Proposition 1.4, there exists a closed interval contained in $[0, 1]$ such that the restriction $\bar{c}$ of c to this interval fulfills the assumption of Proposition 2.4 (read in $T_\eta(\gamma)$ instead of $T_{r_5}(\gamma_0)$). So we get the inequalities $L =$ length of $c \geqslant$ length of $\bar{c} \geqslant \ell =$ length of γ, which proves the first part of the proposition. Moreover, if $L = \ell$, the images of the paths c and $\bar{c}$ must coincide, so $\bar{c}$ shares with γ the same endpoints and the last part of Proposition 2.6 follows from that of Proposition 2.4. □

Corollary 2.7 (each geodesic is uniquely determined by its endpoints). *Take $r_7 > 0$ as in Proposition 2.6. For each $(p, q) \in T_{r_7}(\gamma_0) \times T_{r_7}(\gamma_0)$, there exists at most one unit-speed geodesic of $\gamma : [0, \ell] \to M$ entirely lying in $T_{r_7}(\gamma_0)$ with $\gamma(0) = p$, $\gamma(\ell) = q$.*

Proof. We argue by contradiction. If two distinct unit-speed geodesics of M entirely lying in $T_{r_7}(\gamma_0)$ had the same endpoints, Proposition 2.6 would imply that the length of each geodesic is at least equal to the length of the other; so the geodesics would have equal length. Still by Proposition 2.6, the geodesics would thus coincide, which is absurd. □

3. Proof of Theorem 0.2

Reduction of the proof. We only have to prove the existence of a radius $r > 0$ such that each pair of points of the tube $T_r(\gamma_0)$, located as stated in Theorem 0.2, can be joined by a geodesic with interior lying in $(T_r(\gamma_0))^\circ$. Indeed, suppose we have done so. Then, each such geodesic must be unique (by Corollary 2.7) and minimizing among piecewise C^1 paths sharing the same endpoints and lying in $T_r(\gamma_0)$ (by Proposition 2.6), so the proof is complete.

Strategy. Fixing $p \in T_r(\gamma_0)$, let us consider the subsets

$$\mathfrak{X}_p^+ = \{m \in T_r(\gamma_0), m \neq p, z(m) \geqslant z(p) \text{ and, if } z(p) = 0 \text{ or } \ell_0, z(m) \neq z(p)\},$$
$$\mathfrak{X}_p^- = \{m \in T_r(\gamma_0), m \neq p, z(m) \leqslant z(p) \text{ and, if } z(p) = 0 \text{ or } \ell_0, z(m) \neq z(p)\}.$$

Assuming $z(p) < \ell_0$, we will prove Theorem 0.2 for $q \in \mathfrak{X}_p^+$. Assuming $z(p) > 0$, we would prove it similarly for $q \in \mathfrak{X}_p^-$. Let us proceed to the proof itself. We distinguish two cases.

Case 1: $z(q) - z(p) < \mathfrak{c}/2$. For $\lambda \in [0, 1]$, set $p_\lambda^\perp = [\gamma_0, p](\lambda \mathfrak{r}(p))$ and $q_\lambda^\perp = [\gamma_0, q](\lambda \mathfrak{r}(q))$. Take $r < \mathfrak{c}/2$. Then, for each $\lambda \in [0, 1]$, the points $p_\lambda^\perp$ and $q_\lambda^\perp$ lie in the Riemannian ball $\{m \in M, d(m_0^\perp, m) < \varrho\}$ with center $m_0^\perp = \gamma_0\big(\frac{1}{2}(z(p) + z(q))\big)$ and radius $\varrho = \mathfrak{c}/2 + r < \mathfrak{c}$. Hence there exists a unique minimizing geodesic $c_\lambda : [0, 1] \to M$ going from $p_\lambda^\perp$ to $q_\lambda^\perp$ and such that, for each $t \in [0, 1]$, the map $\lambda \in [0, 1] \to c_\lambda(t) \in M$ is smooth. We must prove that $c_1((0, 1)) \subset (T_r(\gamma_0))^\circ$. To do so, let us argue by connectedness on the set

$$\Lambda = \{\lambda \in [0, 1], c_\lambda((0, 1)) \subset (T_r(\gamma_0))^\circ\}.$$

By construction, Λ is nonempty ($0 \in \Lambda$) and relatively open in $[0, 1]$, so we only have to prove that Λ is closed. Letting $(\lambda_i)_{i \in \mathbb{N}}$ be a sequence of Λ and $\lambda_\infty = \lim_{i \to \infty} \lambda_i \in [0, 1]$, it amounts to prove that $c_{\lambda_\infty}((0, 1)) \subset (T_r(\gamma_0))^\circ$. By continuity, the geodesic c_{λ_∞} ranges in $T_r(\gamma_0)$. If $c_{\lambda_\infty}(\theta) \in C_r(\gamma_0)$ for some $\theta \in (0, 1)$, Proposition 2.2 implies that c_{λ_∞} is constant, so $p_{\lambda_\infty}^\perp = q_{\lambda_\infty}^\perp$. But the latter yields $p = q$, contradicting the assumption $q \in \mathfrak{X}_p^+$.

We are left with ruling out the following property:

$$z(c_{\lambda_\infty}(\theta)) = 0 \text{ or } \ell_0 \quad \text{for some } \theta \in (0, 1). \tag{15}$$

To do so, given $\delta > 0$, we distinguish two subcases as stated in Theorem 0.2.

Subcase 1: $n = 2$. If (15) held, the vector $(dc_{\lambda_\infty}/dt)(\theta)$ would necessarily belong to $\ker dz \setminus \{0\}$. But then, the geodesic $t \mapsto c_{\lambda_\infty}(t)$ would stay for all $t \in [0, 1]$ in the end of the tube given by the equation $z = z(c_{\lambda_\infty}(\theta))$ because, when $n = 2$, the latter is *totally geodesic*. We reach a contradiction, since we have assumed that $z(p) < \ell_0$ and, if $z(p) = 0$, $z(q) \neq 0$.

Subcase 2: $n > 2$ *and either* $|z(p) - z(q)| \geqslant \varsigma$ *or* $\varsigma \leqslant z(p) \leqslant z(q) \leqslant \ell_0 - \varsigma$. If $|z(p) - z(q)| \geqslant \varsigma$, the length ℓ_{λ_∞} of the geodesic c_{λ_∞} must be bounded below by ς due to Lemma 2.3. It follows that $dc^n_{\lambda_\infty}/dt > 0$ if $r > 0$ is taken small enough, due to Proposition 1.2 read with $\delta = \varsigma$. So, in that case, the property (15) cannot hold. If instead $\varsigma \leqslant z(p) \leqslant z(q) \leqslant \ell_0 - \varsigma$, with $|z(p) - z(q)| < \varsigma$, the latter inequality yields $\ell_{\lambda_\infty} \leqslant \varsigma + 2r$, while the former pinching combined with Lemma 2.3 yields $\ell_{\lambda_\infty} \geqslant 2\varsigma$ if (15) holds. In that case, we get the lower bound $r \geqslant \varsigma$ which is absurd, provided $r < \varsigma$. In either case, we conclude that (15) cannot occur for $r > 0$ small enough.

Having proved that $\lambda_\infty \in \Lambda$, we conclude that Λ is closed and hence equal to $[0, 1]$. In particular, $1 \in \Lambda$ so Case 1 is settled.

Case 2: $z(q) - z(p) \geqslant \mathfrak{c}/2$. Here, reading the constant r_1 from Proposition 1.2 with $\delta = \mathfrak{c}/2$, we take $r > 0$ small as done in Proposition 2.5. Furthermore, we consider the subset of the interval $[z(p), \ell_0]$ defined by

$$\mathfrak{Z}^+_p = \big\{\mathfrak{z} \in [z(p), \ell_0], \forall m \in \mathscr{X}^+_p, z(m) = \mathfrak{z} \Longrightarrow T_r(\gamma_0) \text{ is strongly convex for } (p, m)\big\}.$$

By construction, if $\mathfrak{z} \in \mathfrak{Z}^+_p$, the whole interval $[z(p), \mathfrak{z}]$ must lie in $\mathfrak{Z}^+_p$ and, by Case 1, we know that $\mathfrak{Z}^+_p$ contains the interval $\big[z(p), z(p) + \mathfrak{c}/2\big)$. In the next two lemmas, we prove that $\mathfrak{Z}^+_p$ is both closed and relatively open in $[z(p), \ell_0]$. Granted it is, by connectedness, it must coincide with $[z(p), \ell_0]$; hence Theorem 0.2 is established when $z(p) < \ell_0$ and $q \in \mathscr{X}^+_p$. The proof when $z(p) > 0$ and $q \in \mathscr{X}^-_p$ is similar. $\square$

Lemma 3.1. *The subset* $\mathfrak{Z}^+_p$ *is closed.*

Lemma 3.2. *The subset* $\mathfrak{Z}^+_p$ *is relatively open in* $[z(p), \ell_0]$.

Proof of Lemma 3.1. Let $(z_i)_{i \in \mathbb{N}}$ be a sequence of $\mathfrak{Z}^+_p$; set $\mathfrak{z} = \lim_{i\to\infty} z_i \in [z(p), \ell_0]$. We must prove that $\mathfrak{z} \in \mathfrak{Z}^+_p$, so we may assume with no loss of generality that $\mathfrak{z} \geqslant z(p) + \mathfrak{c}/2$. Fix $m \in \mathscr{X}^+_p$ satisfying $z(m) = \mathfrak{z}$ and let $(m_i)_{i\in\mathbb{N}}$ be a sequence of $\mathscr{X}^+_p$ such that, for all $i \in \mathbb{N}$, $z(m_i) = z_i$ and $\lim_{i\to\infty} m_i = m$. For each $i \in \mathbb{N}$, set $t \in [0, 1] \to c_i(t) \in M$ for the unique minimizing geodesic such that $c_i(0) = p$, $c_i(1) = m_i$ and $c_i((0, 1)) \subset (T_r(\gamma_0))^\circ$. By Proposition 1.1, the sequence $\big((dc_i/dt)(0)\big)_{i\in\mathbb{N}}$ is bounded in T_pM; it thus converges toward a vector $V \in T_pM$. By continuity of the map $\exp_p : T_pM \to M$, the geodesic $t \in [0, 1] \to \exp_p(tV) \in M$ (let us denote it by c) satisfies $c(0) = p$, $c(1) = m$ and $c([0, 1]) \subset T_r(\gamma_0)$. For each $t \in (0, 1)$, Proposition 1.2 implies that $z(c(t)) \in (z(p), z(m))$ while, taking $r \leqslant r_4$, we know that $\mathfrak{r}(c(t)) < r$ by Proposition 2.2, so the inclusion $c((0, 1)) \subset (T_r(\gamma_0))^\circ$ must hold. Finally, by Proposition 2.6 and Corollary 2.7, the geodesic c must be minimizing and unique in $T_r(\gamma_0)$. In other words, we have proved that $T_r(\gamma_0)$ is strongly convex for (p, m). Since the point m is arbitrary, we conclude that $\mathfrak{z} \in \mathfrak{Z}^+_p$ as desired. $\square$

Proof of Lemma 3.2. Pick $\mathfrak{z} \in \mathfrak{Z}_p^+$ and $m \in \mathfrak{X}_p^+$ with $z(m) = \mathfrak{z}$. We may take $\mathfrak{z} \in [z(p) + \mathfrak{c}/2, \ell_0)$ without loss, due to Lemma 3.1. Let $t \in [0, 1] \to c_m(t) \in M$ be the geodesic such that $c_m(0) = p$, $c_m(1) = m$ and $c_m((0, 1)) \subset (T_r(\gamma_0))^\circ$. By Proposition 2.5, the Jacobi operator of c_m is negative. Therefore the tangent map $d(\exp_p)\big((dc_m/dt)(0)\big) : T_pM \to T_mM$ is invertible [Aubin 1998, pp. 17–18; do Carmo 1992, pp. 117, 149; Milnor 1963, pp. 98, 100]. The inverse function theorem [Lang 2002] yields a real $\epsilon_m > 0$ such that each point m' lying in the Riemannian ball $B(m, \epsilon_m)$ can be joined to the point p by a unique geodesic $t \in [0, 1] \to c_{m'}(t) = \exp_p(tV') \in M$ with $V' \in T_pM$ close to $V_m = (dc_m/dt)(0)$. Possibly shrinking $\epsilon_m > 0$, we take it such that $z(p) + \mathfrak{c}/4 \leqslant z < \ell_0$ on $\overline{B(m, \epsilon_m)}$. Since the level set $T_r(\gamma_0) \cap \{z = \mathfrak{z}\}$ is compact, it can be covered by the union of finitely many balls $B_i = B(m_i, \epsilon_i)$, $i \in \{1, \dots, N\}$, each constructed like the ball $B(m, \epsilon_m)$. There exists $\theta > 0$ such that the level set $T_r(\gamma_0) \cap \{z = \mathfrak{z} + \theta\}$ remains covered by $\bigcup_{i=1}^N B_i$.

Claim. *The subset $\mathfrak{Z}_p^+$ contains $\mathfrak{z} + \theta$.*

The claim, provisionally taken for granted, implies that $[z(p), \mathfrak{z} + \theta] \subset \mathfrak{Z}_p^+$, so Lemma 3.2, indeed, holds. □

Proof of the claim. Pick $m' \in \mathfrak{X}_p^+$ with $z(m') = \mathfrak{z} + \theta$. There exists $i \in \{1, \dots, N\}$ such that $m' \in B_i$. So $m' = \exp_p(V') \in M$ for a unique vector $V' \in T_pM$ close to $V_i = (dc_{m_i}/dt)(0)$. Moreover, there exists a unique geodesic path $\lambda \in [0, 1] \to \mathfrak{m}(\lambda) \in M$ ranging in B_i such that $\mathfrak{m}(0) = m_i$, $\mathfrak{m}(1) = m'$. Let $\lambda \in [0, 1] \to V_\lambda \in T_pM$ be the corresponding path, derived (like V') from the inverse function theorem as done above, such that $\exp_p(V_\lambda) \equiv \mathfrak{m}(\lambda)$. Set $t \in [0, 1] \to \gamma_\lambda(t) \in M$ for the geodesic path given by $\gamma_\lambda(t) = \exp_p(tV_\lambda)$. From the pinching $z(p) + \mathfrak{c}/4 \leqslant z(\mathfrak{m}(\lambda)) < \ell_0$ combined with Proposition 2.2, we know that $\mathfrak{m}((0, 1)) \subset (T_r(\gamma_0))^\circ$. Let us argue by connectedness on the subset of the interval $[0, 1]$ given by

$$\mathrm{L} = \big\{\lambda \in [0, 1], \gamma_\lambda((0, 1)) \subset (T_r(\gamma_0))^\circ\big\},$$

which is nonempty ($0 \in \mathrm{L}$). The closedness of L can readily be established, arguing as we did for that of $\mathfrak{Z}_p^+$. Let us focus on proving that L is relatively open in $[0, 1]$. If $\lambda \in \mathrm{L}$, the continuity of $\exp_p$ implies the existence of $\mu > 0$ such that $\gamma_{\lambda'}([0, 1]) \subset T_{2r}(\gamma_0)$ for each $\lambda' \in (\lambda - \mu, \lambda + \mu) \cap [0, 1]$. By Lemma 2.3, taking $2r \leqslant r_5$, we know that the length of the geodesic $\gamma_{\lambda'}$ is at least equal to $\mathfrak{c}/4$. By Proposition 1.2 read in $T_{2r}(\gamma_0)$ with $\delta = \mathfrak{c}/4$, we can take $r > 0$ small enough such that $d\gamma_{\lambda'}/dt > 0$; hence $z(\gamma_{\lambda'}((0, 1])) \subset (z(p), \ell_0)$. Furthermore, taking $2r \leqslant r_4$ and applying Proposition 2.2, we get $\mathfrak{r}(\gamma_{\lambda'}(t)) < r$ for $t \in (0, 1)$. It follows that $\lambda' \in \mathrm{L}$; in other words, L is relatively open in $[0, 1]$. By connectedness, we get $\mathrm{L} = [0, 1]$. In particular, $1 \in \mathrm{L}$, from which we readily infer that $m' \in \mathfrak{X}_p^+$. Since m' is arbitrary, we conclude $\mathfrak{z} + \theta \in \mathfrak{Z}_p^+$, as claimed. □

4. Proof of Corollary 0.3

The assumption made in Theorem 0.2 on the geodesic γ_0 is an *open* condition. Given a small real $\varsigma > 0$, we can thus find $r > 0$ such that Theorem 0.2 holds for the geodesic $s \in [-r, \ell_0 + r] \to \gamma_r(s) \in M$ defined as the extension of the geodesic γ_0 to the interval $[-r, \ell_0 + r]$. There still exists a Fermi map about the extended geodesic γ_r; let us stick to our preceding notations for this map. It is important to note the inclusion

$$N_r(\gamma_0) \subset T_r(\gamma_r) \tag{16}$$

which follows from those of $\overline{B(\gamma_0(0), r)}$ and $\overline{B(\gamma_0(\ell_0), r)}$ in $T_r(\gamma_r)$ combined with the identity $N_r(\gamma_0) \equiv T_r(\gamma_0) \cup \overline{B(\gamma_0(0), r)} \cup \overline{B(\gamma_0(\ell_0), r)}$. Given a pair of points (p, q) in $N_r(\gamma_0)$, say, with $z(p) \leqslant z(q)$, we must prove that $N_r(\gamma_0)$ is strongly convex for (p, q). To do so, it suffices to construct a geodesic path from p to q ranging in $(N_r(\gamma_0))^\circ$. Indeed, by (16) combined with Proposition 2.6 and Corollary 2.7 applied in $T_r(\gamma_r)$, such a geodesic path will necessarily be minimizing and unique in $N_r(\gamma_0)$. From Theorem 0.2 applied in $T_r(\gamma_0) \subset N_r(\gamma_0)$, we only have to treat the following two cases.

Case 1: $z(q) - z(p) \geqslant \varsigma$ and either $z(p) < 0$ or $z(q) > \ell_0$. By Theorem 0.2, the tube $T_r(\gamma_r)$ is strongly convex for (p, q). Let $t \in [0, 1] \to \gamma(t) \in M$ denote the geodesic from $\gamma(0) = p$ to $\gamma(1) = q$ such that $\gamma((0, 1)) \subset (T_r(\gamma_r))^\circ$. We must prove that $\gamma((0, 1)) \subset (N_r(\gamma_0))^\circ$. By Proposition 1.2, we know that $d(z \circ \gamma)/dt > 0$ while, by Proposition 2.2, we have $\mathfrak{r} \circ \gamma < r$ on $(0, 1)$. We may assume with no loss of generality the existence of $T \in (0, 1)$ such that either $z(\gamma(T)) = 0$ or $z(\gamma(T)) = \ell_0$. If the former occurs, the restriction of γ to the subinterval $[0, T]$ is minimizing in $T_r(\gamma_r) \cap \{-r \leqslant z \leqslant 0\}$ among piecewise C^1 paths joining p to $\gamma(T)$. Besides, the ball $\overline{B(\gamma_0(0), r)}$ being strongly convex, there exists a unique minimizing geodesic $\tau \in [0, 1] \to c(\tau) \in M$ such that $c(0) = p, c(1) = \gamma(T), c((0, 1]) \subset (B(\gamma_0(0), r))^\circ$. By uniqueness and due to (16), these geodesics must coincide: $c(\tau) \equiv \gamma(\tau T)$. In particular, we do have $\gamma((0, T]) \subset (B(\gamma_0(0), r))^\circ$. Similarly, if the latter occurs, the restriction of γ to the subinterval $[T, 1]$ is minimizing in $T_r(\gamma_r) \cap \{\ell_0 \leqslant z \leqslant \ell_0 + r\}$ among piecewise C^1 paths joining $\gamma(T)$ to q. The ball $\overline{B(\gamma_0(\ell_0), r)}$ being strongly convex, there exists a unique minimizing geodesic $\tau \in [0, 1] \to c(\tau) \in M$ such that $c(0) = \gamma(T)$, $c(1) = q$ and $c([0, 1)) \subset (B(\gamma_0(\ell_0), r))^\circ$. Again, these geodesics must coincide: $c(\tau) \equiv \gamma(\tau + (1 - \tau)T)$. In particular, we do have $\gamma([T, 1)) \subset (B(\gamma_0(\ell_0), r))^\circ$. Case 1 is settled.

Case 2: $z(q) - z(p) < \varsigma$ and either $z(p) < \varsigma$ or $z(q) > \ell_0 - \varsigma$. Here, we may assume that the points p and q lie in the closure of a strongly convex ball B and argue as in Case 1 of the proof of Theorem 0.2, with $T_r(\gamma_0)$ now replaced by $N_r(\gamma_0)$. Doing so, the present proof is reduced to ruling out the analogue of (15), namely the property

$$c_{\lambda_\infty}(\theta) \in \big[\partial B(\gamma_0(0), r) \cap \{z < 0\}\big] \cup \big[\partial B(\gamma_0(\ell_0), r) \cap \{z > \ell_0\}\big] \quad \text{for some } \theta \in (0, 1).$$

This can be done by observing that the geodesic $t \in [0, 1] \to c_{\lambda_\infty}(t) \in M$ is *minimizing* from $p^\perp_{\lambda_\infty}$ to $q^\perp_{\lambda_\infty}$ and by relying on the inclusion (16) combined with the strong convexity of the balls $\overline{B(\gamma_0(0), r)}$ and $\overline{B(\gamma_0(\ell_0), r)}$; we leave it as an exercise.

Acknowledgments

The author is grateful to the referee for valuable comments, some of which led him to clarify the proof of Proposition 1.4.

References

[Aubin 1998] T. Aubin, *Some nonlinear problems in Riemannian geometry*, Springer, Berlin, 1998. MR 99i:58001 Zbl 0896.53003

[do Carmo 1992] M. P. do Carmo, *Riemannian geometry*, Birkhäuser, Boston, 1992. MR 92i:53001 Zbl 0752.53001

[Cheeger and Ebin 2008] J. Cheeger and D. G. Ebin, *Comparison theorems in Riemannian geometry*, AMS Chelsea, Providence, RI, 2008. MR 2009c:53043 Zbl 1142.53003

[Delanoë and Ge 2010] P. Delanoë and Y. Ge, "Regularity of optimal transport on compact, locally nearly spherical, manifolds", *J. Reine Angew. Math.* **646** (2010), 65–115. MR 2012e:49093 Zbl 1200.58025

[Figalli et al. 2012] A. Figalli, L. Rifford, and C. Villani, "Nearly round spheres look convex", *Amer. J. Math.* **134**:1 (2012), 109–139. MR 2876141 Zbl 1241.53031

[Gray 2004] A. Gray, *Tubes*, 2nd ed., Progress in Mathematics **221**, Birkhäuser, Basel, 2004. MR 2004j:53001 Zbl 1048.53040

[Gray and Vanhecke 1982] A. Gray and L. Vanhecke, "The volumes of tubes about curves in a Riemannian manifold", *Proc. London Math. Soc.* (3) **44**:2 (1982), 215–243. MR 83h:53031 Zbl 0491.53035

[Gromov 1981] M. Gromov, *Structures métriques pour les variétés Riemanniennes*, CEDIC/Nathan, Paris, 1981. Translated as *Metric structures for Riemannian and non-Riemannian spaces*, Birkhäuser, Boston, 2007. MR 2007k:53049 Zbl 1113.53001

[Kato 1995] T. Kato, *Perturbation theory for linear operators*, Springer, Berlin, 1995. MR 96a:47025 Zbl 0836.47009

[Klingenberg 1995] W. P. A. Klingenberg, *Riemannian geometry*, 2nd ed., de Gruyter Studies in Mathematics **1**, de Gruyter, Berlin, 1995. MR 95m:53003 Zbl 0911.53022

[Kobayashi and Nomizu 1996] S. Kobayashi and K. Nomizu, *Foundations of differential geometry, I*, Wiley, New York, 1996. MR 97c:53001a Zbl 0119.37502

[Lang 2002] S. Lang, *Introduction to differentiable manifolds*, 2nd ed., Springer, New York, 2002. MR 2003h:58002 Zbl 1008.57001

[Milnor 1963] J. Milnor, *Morse theory*, Annals of Mathematics Studies **51**, Princeton University Press, 1963. MR 29 #634 Zbl 0108.10401

[Protter and Weinberger 1967] M. H. Protter and H. F. Weinberger, *Maximum principles in differential equations*, Prentice-Hall, Englewood Cliffs, NJ, 1967. MR 36 #2935 Zbl 0549.35002

[Spivak 1979] M. Spivak, *A comprehensive introduction to differential geometry, II*, 2nd ed., Publish or Perish, Wilmington, DE, 1979. MR 82g:53003b Zbl 0439.53002

[Whitehead 1932] J. H. C. Whitehead, "Convex regions in the geometry of paths", *Quart. J. Math. Oxford Ser.* **3** (1932), 33–42. Zbl 0004.13102

Received November 22, 2011. Revised May 18, 2013.

PHILIPPE DELANOË
LABORATOIRE J.A. DIEUDONNÉ
UMR 7351 CNRS UNS
UNIVERSITÉ NICE – SOPHIA ANTIPOLIS
FACULTÉ DES SCIENCES
PARC VALROSE
06108 CEDEX 2 NICE
FRANCE

philippe.delanoe@unice.fr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.333

EIGENVALUES OF PERTURBED LAPLACE OPERATORS ON COMPACT MANIFOLDS

ASMA HASSANNEZHAD

We obtain upper bounds for the eigenvalues of the Schrödinger operator $L = \Delta_g + q$ depending on integral quantities of the potential q and a conformal invariant called the *min-conformal volume*. When the Schrödinger operator L is positive, integral quantities of q appearing in upper bounds can be replaced by the mean value of the potential q. The upper bounds we obtain are compatible with the asymptotic behavior of the eigenvalues. We also obtain upper bounds for the eigenvalues of the weighted Laplacian or the Bakry–Émery Laplacian $\Delta_\phi = \Delta_g + \nabla_g \phi \cdot \nabla_g$ using two approaches: first, we use the fact that Δ_ϕ is unitarily equivalent to a Schrödinger operator and we get an upper bound in terms of the L^2-norm of $\nabla_g \phi$ and the min-conformal volume; second, we use its variational characterization and we obtain upper bounds in terms of the L^∞-norm of $\nabla_g \phi$ and a new conformal invariant. The second approach leads to a Buser type upper bound and also gives upper bounds that do not depend on ϕ when the Bakry–Émery Ricci curvature is nonnegative.

1. Introduction and statement of results

We study upper bound estimates for the eigenvalues of Schrödinger operators and weighted Laplace operators or Bakry–Émery Laplace operators.

The author has benefited from the support of the *boursier du gouvernement Français* during her stay in Tours.

MSC2010: 35P15, 47A75, 58J50.

Keywords: Schrödinger operator, Bakry–Émery Laplace operator, eigenvalue, upper bound, conformal invariant.

The Schrödinger operator. Let (M, g) be a compact Riemannian manifold of dimension m and $q \in C^0(M)$. The eigenvalues of the Schrödinger operator $L := \Delta_g + q$ acting on functions constitute a nondecreasing, semibounded sequence of real numbers going to infinity:

$$\lambda_1(\Delta_g + q) \leq \lambda_2(\Delta_g + q) \leq \cdots \leq \lambda_k(\Delta_g + q) \leq \cdots \nearrow \infty.$$

The well-known Weyl law, which describes the asymptotic behavior of the eigenvalues of the Laplacian [Bérard 1986], can be easily extended to the eigenvalues of Schrödinger operators on compact Riemannian manifolds:

$$\lim_{k\to\infty} \lambda_k(\Delta_g + q)\left(\frac{\mu_g(M)}{k}\right)^{2/m} = \alpha_m, \tag{1}$$

where $\alpha_m = 4\pi^2\omega_m^{-2/m}$ and ω_m is the volume of the unit ball in $\mathbb{R}^m$. This says that the normalized eigenvalues $\lambda_k(\Delta_g + q)(\mu_g(M)/k)^{2/m}$ asymptotically tend to a constant depending only on the dimension. However, upper bounds of normalized eigenvalues in general cannot be independent of geometric invariants and the potential q; see [Colbois and Dodziuk 1994] or the introduction of [Hassannezhad 2011]. We shall obtain upper bounds for normalized eigenvalues depending on some geometric invariants and integral quantities of the potential q. These upper bounds are compatible with the asymptotic behavior in (1); that is, they tend asymptotically to a constant depending only on the dimension as k goes to infinity.

Numerous articles have studied how the eigenvalues of L can be controlled in terms of geometric invariants of the manifold and quantities depending on the potential. From the variational characterization of eigenvalues, it is easy to see that

$$\lambda_1(\Delta_g + q) \leq \frac{1}{\mu_g(M)} \int_M q \, d\mu_g.$$

For the second eigenvalue $\lambda_2(\Delta_g + q)$, El Soufi and Ilias [1992, Theorem 2.2] obtained an upper bound in terms of the mean value of the potential q and a conformal invariant:

$$\lambda_2(\Delta_g + q) \leq m\left(\frac{V_c([g])}{\mu_g(M)}\right)^{2/m} + \frac{\int_M q \, d\mu_g}{\mu_g(M)}, \tag{2}$$

where $V_c([g])$ is the conformal volume defined by Li and Yau [1982] which only depends on the conformal class of g, denoted by $[g]$.

For a compact orientable Riemannian surface (Σ_γ, g) of genus γ, as a consequence of inequality (2), they obtained the following inequality, where $\lfloor\ \rfloor$ denotes the floor function:

$$\lambda_2(\Delta_g + q) \leq \frac{8\pi}{\mu_g(\Sigma_\gamma)}\left\lfloor \frac{\gamma + 3}{2} \right\rfloor + \frac{\int_{\Sigma_\gamma} q \, d\mu_g}{\mu_g(\Sigma_\gamma)}. \tag{3}$$

For higher eigenvalues of Schrödinger operators, Grigor'yan, Netrusov and Yau [Grigor'yan et al. 2004] proved a general and abstract result that can be stated in the case of Schrödinger operators as follows. Given positive constants N and C_0, assume that a compact Riemannian manifold (M, g) has the $(2, N)$-covering property (that is, each ball of radius r can be covered by N balls of radius $r/2$) and $\mu_g(B(x, r)) \leq C_0 r^2$ for every $x \in M$ and every $r > 0$. Then, for every $q \in C^0(M)$, we have (see [Grigor'yan et al. 2004, Theorem 1.2 (1.14)])

$$\lambda_k(\Delta_g + q) \leq \frac{Ck + \delta^{-1} \int_M q^+ \, d\mu_g - \delta \int_M q^- \, d\mu_g}{\mu_g(M)}, \tag{4}$$

where $\delta \in (0, 1)$ is a constant which depends only on N, $C > 0$ is a constant which depends on N and C_0, and $q^{\pm} = \max\{|\pm q|, 0\}$.

Moreover, if L is a positive operator, then we have (see [Grigor'yan et al. 2004, Theorem 5.15])

$$\lambda_k(\Delta_g + q) \leq \frac{Ck + \int_M q \, d\mu_g}{\epsilon \mu_g(M)}, \tag{5}$$

where $\epsilon \in (0, 1)$ depends only on N and C depends on N and C_0.

The above inequalities in dimension two have a special feature as follows. Let Σ_γ be a compact orientable Riemannian surface of genus γ. Then, for every Riemannian metric g on Σ_γ and every $q \in C^0(\Sigma_\gamma)$, we have (see [Grigor'yan et al. 2004, Theorem 5.4])

$$\lambda_k(\Delta_g + q) \leq \frac{Q(\gamma + 1)k + \delta^{-1} \int_{\Sigma_\gamma} q^+ d\mu_g - \delta \int_{\Sigma_\gamma} q^- d\mu_g}{\mu_g(\Sigma_\gamma)},$$

where $\delta \in (0, 1)$ and $Q > 0$ are absolute constants.

Inequalities (4) and (5) are not compatible with the asymptotic behavior regarding the power of k, except in dimension two. Yet, for surfaces, the limit of the above upper bound for normalized eigenvalues depends on the genus γ as k goes to infinity. Therefore, it is not compatible with (1).

We obtain upper bounds which generalize and improve the above inequalities without imposing any condition on the metric and which are compatible with the asymptotic behavior. Before stating our theorem, we need to recall the definition of the *min-conformal volume*. For a compact Riemannian manifold (M, g), its *min-conformal volume* is defined as follows (see [Hassannezhad 2011]):

$$V([g]) = \inf\{\mu_{g_0}(M) : g_0 \in [g], \ \mathrm{Ricci}_{g_0} \geq -(m-1)\}.$$

Theorem 1.1. *There exist positive constants $\alpha_m \in (0, 1)$, B_m, and C_m depending only on m such that, for every compact m-dimensional Riemannian manifold (M, g), every potential $q \in C^0(M)$, and every $k \in \mathbb{N}^*$, we have*

$$\text{(6)}\quad \lambda_k(\Delta_g+q) \le \frac{\alpha_m^{-1}\int_M q^+\,d\mu_g - \alpha_m\int_M q^-\,d\mu_g}{\mu_g(M)} + B_m\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + C_m\left(\frac{k}{\mu_g(M)}\right)^{2/m}.$$

In particular, when the potential q is nonnegative, one has

$$\text{(7)}\qquad \lambda_k(\Delta_g+q) \le A_m\frac{\int_M q\,d\mu_g}{\mu_g(M)} + B_m\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + C_m\left(\frac{k}{\mu_g(M)}\right)^{2/m},$$

where $A_m = \alpha_m^{-1}$.

We also obtain upper bounds for eigenvalues of positive Schrödinger operators. Note that the positivity of the Schrödinger operator $L = \Delta_g + q$ implies that $\int_M q$ is nonnegative, and q here may not be nonnegative. The following upper bound generalizes inequalities (5) and (7).

Theorem 1.2. *There exist constants $A_m > 1$, B_m, and C_m depending only on m such that if $L = \Delta_g + q$ with $q \in C^0(M)$ is a positive operator, then, for every compact m-dimensional Riemannian manifold (M^m, g) and every $k \in \mathbb{N}^*$, we have*

$$\lambda_k(\Delta_g+q) \le A_m\frac{\int_M q\,d\mu_g}{\mu_g(M)} + B_m\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + C_m\left(\frac{k}{\mu_g(M)}\right)^{2/m}.$$

Given the Schrödinger operator $L = \Delta_g + q$, for every $\varepsilon > 0$, the Schrödinger operator $\widetilde{L} = \Delta_g + q - \lambda_1(L) + \varepsilon$ is positive and $\lambda_k(\widetilde{L}) = \lambda_k(L) - \lambda_1(L) + \varepsilon$. When ε goes to zero, Theorem 1.1 leads to the following.

Corollary 1.3. *Under the assumptions of Theorem 1.1, we get*

$$\lambda_k(\Delta_g+q) \le A_m\frac{\int_M q\,d\mu_g}{\mu_g(M)} + (1-A_m)\lambda_1(\Delta_g+q) + B_m\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + C_m\left(\frac{k}{\mu_g(M)}\right)^{2/m}.$$

In the two-dimensional case, for a compact orientable Riemannian surface (Σ_γ, g) of genus γ, thanks to the uniformization and Gauss–Bonnet theorems, one has $V([g]) \le 4\pi\gamma$. Therefore, in compact orientable Riemannian surfaces, one can replace the min-conformal volume by the topological invariant $4\pi\gamma$ in the above inequalities.

Corollary 1.4. *There exist absolute constants $a \in (0, 1)$, A, and B such that, for every compact orientable Riemannian surface (Σ_γ, g) of genus γ, every potential $q \in C^0(M)$, and every $k \in \mathbb{N}^*$, we have*

$$\text{(8)}\qquad \lambda_k(\Delta_g+q)\mu_g(\Sigma_\gamma) \le \int_{\Sigma_\gamma}(aq^+ - a^{-1}q^-)\,d\mu_g + A\gamma + Bk.$$

And if L is a positive operator,

$$\lambda_k(\Delta_g + q)\mu_g(\Sigma_\gamma) \le a \int_{\Sigma_\gamma} q \, d\mu_g + A\gamma + Bk.$$

An interesting application of Theorem 1.1 is the case of weighted Laplace operators or Bakry–Émery Laplace operators.

Bakry–Émery Laplacian. Let (M, g) be a Riemannian manifold and $\phi \in C^2(M)$. The corresponding weighted Laplace operator Δ_ϕ is defined by

$$\Delta_\phi = \Delta_g + \nabla_g \phi \cdot \nabla_g .$$

This operator is associated with the quadratic functional $\int_M |\nabla_g f|^2 e^{-\phi} \, d\mu_g$, that is,

$$\int_M \Delta_\phi f h e^{-\phi} \, d\mu_g = \int_M \langle \nabla_g f, \nabla_g h \rangle e^{-\phi} \, d\mu_g .$$

This operator is an elliptic operator on $C_c^\infty(M) \subseteq L^2(e^{-\phi} \, d\mu_g)$ and can be extended to a selfadjoint operator with the weighted measure $e^{-\phi} \, d\mu_g$. In this sense, it arises as a generalization of the Laplacian. The weighted Laplace operator Δ_ϕ is also known as the diffusion operator or the Bakry–Émery Laplace operator which is used to study the diffusion process; see, for instance, the pioneering work of Bakry and Émery [1985] or [Lott 2007; Lott and Villani 2009]. The triple (M, g, ϕ) is called a Bakry–Émery manifold, where $\phi \in C^2(M)$ and (M, g) is a Riemannian manifold with the weighted measure $e^{-\phi} d\mu_g$; see [Lu and Rowlett 2012; Rowlett 2010]. The interplay between the geometry of M and the behavior of ϕ is mostly taken into account by means of a new notion of curvature called the Bakry–Émery Ricci tensor[1], which is defined by

$$\mathrm{Ricci}_\phi = \mathrm{Ricci}_g + \mathrm{Hess}\, \phi .$$

Our aim is to find upper bounds for the eigenvalues of Δ_ϕ denoted by $\lambda_k(\Delta_\phi)$ in terms of the geometry of M and of properties of ϕ.

Upper bounds for the first eigenvalue $\lambda_1(\Delta_\phi)$ of complete noncompact Riemannian manifolds have been recently considered in several works; see [Munteanu and Wang 2012; Setti 1998; Su and Zhang 2012; Wu 2010; 2012]. These upper bounds depend on the L^∞-norm of $\nabla_g \phi$ and a lower bound of the Bakry–Émery Ricci tensor.

Let (M, g, ϕ) be a complete noncompact Bakry–Émery manifold of dimension m with $\mathrm{Ricci}_\phi \ge -\kappa^2(m-1)$ and $|\nabla_g \phi| \le \sigma$ for some constants $\kappa \ge 0$ and $\sigma > 0$. Then we have, by [Su and Zhang 2012, Proposition 2.1] (see also [Munteanu and

[1] The Bakry–Émery Ricci tensor Ricci_ϕ is also referred to as the ∞-Bakry–Émery Ricci tensor. We denote Ricci_ϕ and $\mathrm{Hess}\, \phi$ by $\mathrm{Ricci}_\phi(M, g)$ and $\mathrm{Hess}_g\, \phi$ wherever any confusion might occur.

Wang 2012; Wu 2010; 2012])

$$\lambda_1(\Delta_\phi) \le \tfrac{1}{4}((m-1)\kappa+\sigma)^2. \tag{9}$$

In particular, if $\mathrm{Ricci}_\phi \ge 0$, we have

$$\lambda_1(\Delta_\phi) \le \tfrac{1}{4}\sigma^2. \tag{10}$$

We consider compact Bakry–Émery manifolds and we present two approaches to obtain upper bounds for the eigenvalues of the Bakry–Émery Laplace operator in terms of the geometry of M and the properties of ϕ.

First approach. One can see that Δ_ϕ is unitarily equivalent to the Schrödinger operator $L=\Delta_g+\frac{1}{2}\Delta_g\phi+\frac{1}{4}|\nabla_g\phi|^2$; see, for example, [Setti 1998, p. 28]. Therefore, as a consequence of Theorem 1.2, we obtain an upper bound for $\lambda_k(\Delta_\phi)$ in terms of the min-conformal volume and the L^2-norm of $\nabla_g\phi$.

Theorem 1.5. *There exist constants A_m, B_m, and C_m depending on $m \in \mathbb{N}^*$, such that, for every m-dimensional compact Bakry–Émery manifold (M, g, ϕ), we have*

$$\lambda_k(\Delta_\phi) \le A_m \frac{1}{\mu_g(M)}\|\nabla_g\phi\|^2_{L^2(M)} + B_m\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + C_m\left(\frac{k}{\mu_g(M)}\right)^{2/m}.$$

It is worth noticing that in full generality it is not possible to obtain upper bounds which do not depend on ϕ; see, for instance, [Su and Zhang 2012, Section 2]. However, we will see that for compact manifolds with nonnegative Bakry–Émery Ricci curvature we can find upper bounds which do not depend on ϕ (see Corollary 1.8).

In the two-dimensional case, as a result of Corollary 1.4, we obtain the following.

Corollary 1.6. *There exist absolute constants $a \in (0, 1)$, A, and B such that, for every compact orientable Riemannian surface (Σ_γ, g) of genus γ and every $k \in \mathbb{N}^*$,*

$$\lambda_k(\Delta_\phi)\mu_g(\Sigma_\gamma) \le a\|\nabla_g\phi\|^2_{L^2(\Sigma_\gamma)} + A\gamma + Bk.$$

Second approach. This approach is based on using the technique introduced in [Hassannezhad 2011], which was successfully applied for the Laplace operator Δ_g on Riemannian manifolds [Hassannezhad 2011, Theorem 1.1]. We obtain upper bounds for eigenvalues of Δ_ϕ in terms of a conformal invariant. We also obtain a Buser type upper bound for $\lambda_k(\Delta_\phi)$ (see Corollary 1.9).

Definition 1.1. Let (M, g, ϕ) be a compact Bakry–Émery manifold. We define the ϕ-min-conformal volume as

$$V_\phi([g]) = \inf\{\mu_\phi(M, g_0) : g_0 \in [g], \mathrm{Ricci}_\phi(M, g_0) \ge -(m-1)\}, \tag{11}$$

where $\mu_\phi(M, g_0)$ is the weighted measure[2] of M with respect to the metric g_0.

[2]For a Bakry–Émery manifold (M, g, ϕ), when μ_ϕ is the weighted measure with respect to the

Note that up to dilations[3] there is always a Riemannian metric $g_0 \in [g]$ such that $\mathrm{Ricci}_\phi(M, g_0) \geq -(m-1)$. We are now ready to state our theorem.

Theorem 1.7. *There exist positive constants $A(m)$ and $B(m)$ depending only on $m \in \mathbb{N}^*$ such that, for every compact Bakry–Émery manifold (M, g, ϕ) with $|\nabla_g \phi| \leq \sigma$, for some $\sigma \geq 0$ and for every $k \in \mathbb{N}^*$, we have*

$$\lambda_k(\Delta_\phi) \leq A(m)\max\{\sigma^2, 1\}\left(\frac{V_\phi([g])}{\mu_\phi(M)}\right)^{2/m} + B(m)\left(\frac{k}{\mu_\phi(M)}\right)^{2/m}. \tag{12}$$

If a metric g is conformally equivalent to a metric g_0 with $\mathrm{Ricci}_\phi(M, g_0) \geq 0$, $V_\phi([g]) = 0$. Thus an immediate consequence of Theorem 1.7 is the following.

Corollary 1.8. *There exists a positive constant $A(m)$ depending only on $m \in \mathbb{N}^*$ such that, for every compact Bakry–Émery manifold (M, g, ϕ) with $V_\phi([g]) = 0$ and for every $k \in \mathbb{N}^*$,*

$$\lambda_k(\Delta_\phi) \leq A(m)\left(\frac{k}{\mu_\phi(M)}\right)^{2/m}. \tag{13}$$

The above upper bound is similar to the upper bound for the eigenvalues of the Laplacian in Riemannian manifolds (M, g) when $V([g]) = 0$; see [Korevaar 1993].

If $\mathrm{Ricci}_\phi(M) \geq -\kappa^2(m-1)$ for some $\kappa \geq 0$, then, for $g_0 = \kappa^2 g$, one has $\mathrm{Ricci}_\phi(M, g_0) \geq -(m-1)$ and $V_\phi([g]) \leq \mu_\phi(M, g_0) = \kappa^m \mu_\phi(M, g)$. Replacing in inequality (12), we get a Buser type upper bound for the eigenvalues of the Bakry–Émery Laplacian.

Corollary 1.9 (Buser type upper bound). *There are positive constants $A(m)$ and $B(m)$ depending only on $m \in \mathbb{N}^*$ such that, for every compact Bakry–Émery manifold (M, g, ϕ) with $\mathrm{Ricci}_\phi(M) > -\kappa^2(m-1)$ and $|\nabla_g \phi| \leq \sigma$ for some $\kappa \geq 0$ and $\sigma \geq 0$, and for every $k \in \mathbb{N}^*$, we have*

$$\lambda_k(\Delta_\phi) \leq A(m)\max\{\sigma^2, 1\}\kappa^2 + B(m)\left(\frac{k}{\mu_\phi(M)}\right)^{2/m}.$$

A weaker version of Corollary 1.9 can be proved directly by the classic idea used by Buser [1979] and Li and Yau [1980]. We refer the reader to the appendix, where we give a simple and direct proof.

Remark 1.1. All of the results mentioned above for compact manifolds are also valid when one considers bounded subdomains of complete manifolds with the Neumann boundary condition.

---

metric g, we simply denote the weighted measure of a measurable subset A of M by $\mu_\phi(A)$ instead of $\mu_\phi(A, g)$.

[3]Note that $\mathrm{Hess}\,\phi$ and Ricci_g do not change under dilations. If $\mathrm{Ricci}_\phi(M, g) \geq -\kappa^2(m-1)g$, for all $\alpha > 0$, $\mathrm{Ricci}_\phi(M, g_0) := \mathrm{Ricci}_\phi(M, \alpha g) = \mathrm{Ricci}_\phi(M, g) \geq -\kappa^2(m-1)g = -(\kappa^2/\alpha)(m-1)g_0$.

2. Preliminaries and technical tools

Basic definitions. A *capacitor* is a pair of Borel sets (F, G) in a topological space satisfying $F \subsetneq G$.

We say that a metric space (X, d) satisfies the $(\kappa, N; \rho)$-*covering property* if each ball of radius $0 < r \le \rho$ can be covered by N balls of radius r/κ. We sometimes call this the local covering property when $\rho < \infty$.

For any $x \in X$ and $0 \le r \le R$, we define the annulus $A(x, r, R)$ as

$$A(x, r, R) := B(x, R) \setminus B(x, r) = \{y \in X : r \le d(x, y) < R\}.$$

Note that $A(x, 0, R) = B(x, R)$. If $F = A(x, r, R)$ and $\lambda \ge 1$, we define ${}^{\lambda}F := A(x, \lambda^{-1}r, \lambda R)$. For $F \subseteq X$ and $r > 0$, we denote by F^r the r-neighborhood of F:

$$F^r = \{x \in X : d(x, F) \le r\}.$$

Here we state the key method that we use in order to obtain our results. This method was introduced in [Hassannezhad 2011] and was inspired by two elaborate constructions given in [Colbois and Maerten 2008; Grigor'yan et al. 2004]. It leads to the construction of a "nice" family of capacitors, crucial to estimating the eigenvalues of Schrödinger operators and Bakry–Émery operators via capacities.

Capacity on Riemannian manifolds. For each capacitor (F, G) in a Riemannian manifold (M, g) of dimension m, we define the capacity and the m-capacity by

$$\text{(14)} \quad \operatorname{cap}_g(F,G) = \inf_{\varphi\in\mathcal{T}} \int_M |\nabla_g\varphi|^2\, d\mu_g \quad \text{and} \quad \operatorname{cap}^{(m)}_{[g]}(F,G) = \inf_{\varphi\in\mathcal{T}} \int_M |\nabla_g\varphi|^m\, d\mu_g,$$

respectively, where $\mathcal{T} = \mathcal{T}(F, G)$ is the set of all functions $\varphi \in C_0^\infty(M)$ such that $\operatorname{supp}\varphi \subset G$, $0 \le \phi \le 1$, and $\varphi \equiv 1$ in a neighborhood of F. If $\mathcal{T}(F, G)$ is empty, $\operatorname{cap}_g(F, G) = \operatorname{cap}^{(m)}_{[g]}(F, G) = +\infty$.

Proposition 2.1 ([Hassannezhad 2012, Theorem 1.2.1]; see also [Hassannezhad 2011]). *Let (X, d, μ) be a metric measure space with a nonatomic Borel measure μ satisfying the $(2, N; \rho)$-covering property. Then, for every $n \in \mathbb{N}^*$, there exists a family of capacitors $\mathcal{A} = \{(F_i, G_i)\}_{i=1}^n$ with the following properties*:

(i) $\mu(F_i) \ge \nu := \mu(X)/(8c^2 n)$, *where c is a constant depending only on N.*

(ii) *The G_i are mutually disjoint.*

(iii) *The family $\mathcal{A}$ is such that either*

  (a) *all the F_i are annuli with outer radii smaller than ρ and $G_i = {}^2F_i$, or*

  (b) *all the F_i are domains in X and $G_i = F_i^{r_0}$ with $r_0 = \frac{1}{1600}\rho$.*

As a consequence of this proposition, we have:

Lemma 2.2. *Let (M^m, g, μ) be a compact Riemannian manifold with a nonatomic Borel measure μ. Then there exist positive constants $c(m) \in (0, 1)$ and $\alpha(m)$ depending only on the dimension such that, for every $k \in \mathbb{N}^*$, there exists a family $\{(F_i, G_i)\}_{i=1}^k$ of mutually disjoint capacitors with the following properties*:

(I) $\mu(F_i) > c(m)\dfrac{\mu(M)}{k}$.

(II) $\operatorname{cap}_g(F_i, G_i) \leq \dfrac{\mu_g(M)}{k}\left[\dfrac{1}{r_0^2}\left(\dfrac{V([g])}{\mu_g(M)}\right)^{2/m} + \alpha(m)\left(\dfrac{k}{\mu_g(M)}\right)^{2/m}\right]$, *with* $r_0 = \frac{1}{1600}$.

Proof of Lemma 2.2. Take the metric measure space (M, d_{g_0}, μ), where $g_0 \in [g]$ with $\operatorname{Ricci}_{g_0} \geq -(m-1)$ and d_{g_0} is the distance associated to the Riemannian metric g_0. It is easy to verify that (M, d_{g_0}, μ) has the $(2, N; 1)$-covering property, where N is a constant depending only on the dimension [Hassannezhad 2011]. Therefore, Proposition 2.1 implies that for every $k \in \mathbb{N}^*$ there is a family of $3k$ mutually disjoint capacitors $\{(F_i, G_i)\}_{i=1}^{3k}$ satisfying the following properties (see [Grigor'yan et al. 2004, Proposition 3.1] for more justification):

- $\mu(F_i) > c(m)\mu(M)/k$, where $c(m) \in (0, 1)$ is a positive constant depending only on the dimension.
- Either
  - (a) all the F_i are annuli with outer radii smaller than 1 and $\operatorname{cap}_{[g]}^{(m)}(F_i, {}^2F_i) \leq Q_m$, where the constant Q_m depends only on the dimension, and $G_i = {}^2F_i$;

  or
  - (b) all the F_i are domains in M and $G_i = F_i^{r_0}$, where $r_0 = \frac{1}{1600}$.

Hence, the family of $\{(F_i, G_i)\}_{i=1}^{3k}$ has property (I). We now show that at least k of the capacitors satisfy property (II). We first find an upper bound for the m-capacity $\operatorname{cap}_{[g]}^{(m)}(F_i, G_i)$. If all the F_i are annuli, we already have an estimate by property (a). If the F_i are domains, one can define a family of functions $\varphi_i \in \mathcal{T}(F_i, G_i)$, $1 \leq i \leq 3k$, such that $|\nabla_{g_0}\varphi_i| \leq 1/r_0$. Then

$$\operatorname{cap}_{[g]}^{(m)}(F_i, G_i) \leq \int_M |\nabla_{g_0}\varphi_i|^m \, d\mu_{g_0} \leq \frac{1}{r_0^m}\mu_{g_0}(G_i).$$

Since $G_1, \dots, G_{3k}$ are mutually disjoint, there exist at least $2k$ of them so that $\mu_{g_0}(G_i) \leq \mu_{g_0}(M)/k$. Similarly, there exist at least $2k$ sets (not necessarily the same ones) such that $\mu_g(G_i) \leq \mu_g(M)/k$. Therefore, up to reordering, we assume that the first k of them (that is, $G_1, \dots, G_k$) satisfy the inequalities

$$\mu_g(G_i) \leq \mu_g(M)/k \quad \text{and} \quad \mu_{g_0}(G_i) \leq \mu_{g_0}(M)/k.$$

Hence, in general, there exist k capacitors (F_i, G_i), $1 \le i \le k$, with

$$\mathrm{cap}_{[g]}^{(m)}(F_i, G_i) \le Q_m + \frac{1}{r_0^m}\frac{\mu_{g_0}(M)}{k}.$$

The left side of this inequality is a conformal invariant. Now, taking the infimum over $g_0 \in [g]$ with $\mathrm{Ricci}_{g_0} \ge -(m-1)$, we get

$$\mathrm{cap}_{[g]}^{(m)}(F_i, G_i) \le Q_m + \frac{1}{r_0^m}\frac{V([g])}{k}.$$

Now, for every $\varepsilon > 0$, we consider plateau functions $\{f_i\}_{i=1}^k$, $f_i \in \mathcal{T}(F_i, G_i)$, with

$$\int_M |\nabla_g f_i|^m \, d\mu_g \le \mathrm{cap}_{[g]}^{(m)}(F_i, G_i) + \varepsilon.$$

Therefore,

$$\begin{aligned}
(15)\quad \mathrm{cap}_g(F_i, G_i) &\le \int_M |\nabla_g f_i|^2 d\mu_g \le \left(\int_M |\nabla_g f_i|^m d\mu_g\right)^{2/m} \left(\int_M 1_{\mathrm{supp}\, f_i}\, d\mu_g\right)^{1-2/m} \\
&\le \left(\mathrm{cap}_{[g]}^{(m)}(F_i, G_i) + \varepsilon\right)^{2/m} \mu_g(G_i)^{1-2/m} \\
&\le \left(Q_m + \frac{1}{r_0^m}\frac{V([g])}{k} + \varepsilon\right)^{2/m} \mu_g(G_i)^{1-2/m} \\
&\le \left[Q_m^{2/m} + \frac{1}{r_0^2}\left(\frac{V([g])}{k}\right)^{2/m} + \varepsilon^{2/m}\right]\left(\frac{\mu_g(M)}{k}\right)^{1-2/m}.
\end{aligned}$$

where the last inequality is due to the well-known fact that

$$(a+b)^s \le a^s + b^s$$

when a, b are nonnegative real numbers and $0 < s \le 1$. Letting ε tend to zero, we obtain property (II). This completes the proof. □

Capacity on Bakry–Émery manifolds. In an analogous way, we define the capacity in a Bakry–Émery manifold (M, g, ϕ). For each capacitor (F, G) in a Bakry–Émery manifold (M, g, ϕ) of dimension m, the capacity and the m-capacity are defined as

$$(16)\quad \mathrm{cap}_\phi(F, G) = \inf_{\varphi \in \mathcal{T}} \int_M |\nabla_g \varphi|^2 \, d\mu_\phi \quad \text{and} \quad \mathrm{cap}_\phi^{(m)}(F, G) = \inf_{\varphi \in \mathcal{T}} \int_M |\nabla_g \varphi|^m \, d\mu_\phi,$$

respectively, where $\mathcal{T} = \mathcal{T}(F, G)$ is the set of all functions $\varphi \in C_0^\infty(M)$ such that $\mathrm{supp}\, \varphi \subset G$, $0 \le \phi \le 1$ and $\varphi \equiv 1$ in a neighborhood of F. If $\mathcal{T}(F, G)$ is empty, $\mathrm{cap}_\phi(F, G) = \mathrm{cap}_\phi^{(m)}(F, G) = +\infty$.

We prove a similar lemma below (Lemma 2.2). We first show that every compact Bakry–Émery manifold satisfies the assumptions of Proposition 2.1. Thanks to a volume comparison theorem for Bakry–Émery manifolds, which we quote next, we can show that such have the local covering property (see Lemma 2.4).

Theorem 2.3 (volume comparison theorem [Wei and Wylie 2009]). *Let* (M, g, ϕ) *be a compact Bakry–Émery manifold with* $Ricci_\phi \geq \alpha(m-1)$. *If* $\partial_r \phi \geq -\sigma$ *with respect to geodesic polar coordinates centered at* x, *then, for every* $0 < r \leq R$, *we have* (*assume* $R \leq \pi/2\sqrt{\alpha}$ *if* $\alpha > 0$)

$$\frac{\mu_\phi(B(x,R))}{\mu_\phi(B(x,r))} \leq e^{\sigma R} \frac{v(m,R,\alpha)}{v(m,r,\alpha)}, \tag{17}$$

and, *in particular*, *letting* r *tend to zero yields*

$$\mu_\phi(B(x,R)) \leq e^{\sigma R} v(m,R,\alpha), \tag{18}$$

where $v(m,r,\alpha)$ *is the volume of a ball of radius* r *in the simply connected space form of constant sectional curvature* α.

Lemma 2.4. *Let* (M, g, ϕ) *be a compact Bakry–Émery manifold with* $Ricci_\phi \geq -\kappa^2(m-1)$ *and* $|\nabla_g \phi| \leq \sigma$ *for some* $\kappa \geq 0$ *and* $\sigma \geq 0$. *There exist constants* $N(m) \in \mathbb{N}^*$ *and* $\xi = \xi(\sigma,\kappa) > 0$ *such that* (M, g, ϕ) *satisfies the* $(2, N; \xi)$*-covering property. Moreover*, *there exists a positive constant* $C(m)$ *such that*, *for every* $0 \leq r < R \leq \xi$ *and* $x \in M$, *the annulus* $A = A(x, r, R)$ *satisfies* $\mathrm{cap}_\phi^{(m)}(A, {}^2A)) \leq C(m)$.

Proof. Take $\xi = \min\{1/\sigma, 1/\kappa\}$. (Take $\xi = \infty$ if $\sigma = \kappa = 0$.) We first show that (M, μ_ϕ) has the doubling property for $r < 4\xi$, that is,

$$\mu_\phi(B(x,r)) \leq c\mu_\phi(B(x,r/2)), \quad 0 < r < 4\xi,$$

for some positive constant c. From this, it is easy to deduce that (M, μ_ϕ) has the $(2, N; \xi)$-covering property, for example with $N = c^4$. To prove the doubling property, according to inequality (17) we have

$$\frac{\mu_\phi(B(x,r))}{\mu_\phi(B(x,r/2))} \leq e^{\sigma r} \frac{v(m,r,-\kappa^2)}{v(m,r/2,-\kappa^2)} = e^{\sigma r} \frac{v(m,\kappa r,-1)}{v(m,\kappa r/2,-1)}.$$

Take $\tilde{r} := \kappa r$. Then, for $0 < r < 4\xi = 4\min\{1/\sigma, 1/\kappa\}$, we have

$$e^{\sigma r} \frac{v(m,\kappa r,-1)}{v(m,\kappa r/2,-1)} \leq e^4 \frac{v(m,\tilde{r},-1)}{v(m,\tilde{r}/2,-1)} \leq c(m),$$

where

$$c(m) := \sup_{\tilde{r} \in (0,4)} e^4 \frac{v(m,\tilde{r},-1)}{v(m,\tilde{r}/2,-1)}.$$

Thus

$$\frac{\mu_\phi(B(x,r))}{\mu_\phi(B(x,r/2))} \leq c(m) \quad \text{for every } 0 < r < \xi.$$

Therefore, (M, g, ϕ) has the $(2, N; \xi)$-covering property for $N = c^4(m)$.

To estimate the capacity of an annulus, we now follow the same argument as in [Hassannezhad 2011, p. 3430]. Let $A = A(x, r, R)$ and let $f \in \mathcal{T}(A, {}^2A)$ be given by

$$f(y) = \begin{cases} 1 & \text{if } y \in A(x, r, R), \\ 2d_{g_0}(y, B(x, r/2))/r & \text{if } y \in A(x, r/2, r) \text{ and } r \neq 0, \\ 1 - d_{g_0}(y, B(x, R))/R & \text{if } y \in A(x, R, 2R), \\ 0 & \text{if } y \in M \setminus A(x, r/2, 2R). \end{cases} \tag{19}$$

We have

$$|\nabla_{g_0} f| \leq \begin{cases} 2/r & \text{on } B(x, r) \setminus B(x, r/2), \\ 1/R & \text{on } B(x, 2R) \setminus B(x, R). \end{cases}$$

Therefore,

$$\begin{aligned} \operatorname{cap}_\phi^{(m)}(A, {}^2A) &\leq \int_M |\nabla_g f|^m d\mu_\phi \leq \left(\frac{2}{r}\right)^m \mu_\phi(A(x, r/2, r)) + \left(\frac{1}{R}\right)^m \mu_\phi(A(x, R, 2R)) \\ &\leq \left(\frac{2}{r}\right)^m \mu_\phi(B(x, r)) + \left(\frac{1}{R}\right)^m \mu_\phi(B(x, 2R)). \end{aligned}$$

Having inequality (18), we get

$$\begin{aligned} \operatorname{cap}_\phi^{(m)}(A, {}^2A) &\leq \left(\frac{2}{r}\right)^m e^{\sigma r} v(m, r, -\kappa^2) + \left(\frac{1}{R}\right)^m e^{2\sigma R} v(m, 2R, -\kappa^2) \\ &= \left(\frac{2}{\kappa r}\right)^m e^{\sigma r} v(m, \kappa r, -1) + \left(\frac{1}{\kappa R}\right)^m e^{2\sigma R} v(m, 2\kappa R, -1). \end{aligned}$$

Take $\tilde{r} := \kappa r$ and $\widetilde{R} := \kappa R$. Then, for every $0 < r < R \leq 2\xi = 2\min\{1/\sigma, 1/\kappa\}$, we get

$$\operatorname{cap}_\phi^{(m)}(A, {}^2A) \leq \left(\frac{2}{\tilde{r}}\right)^m e^2 v(m, \tilde{r}, -1) + \left(\frac{1}{\widetilde{R}}\right)^m e^4 v(m, 2\widetilde{R}, -1). \tag{20}$$

Setting $C(m)$ to the supremum of the expression on the right side over $\tilde{r}, \widetilde{R} \in (0, 2)$ completes the proof. $\square$

Lemma 2.5. *Let (M^m, g, ϕ) be a compact Bakry–Émery manifold with $|\nabla_g \phi| \leq \sigma$ for some $\sigma \geq 0$. There exist positive constants $c(m) \in (0, 1)$ and $\alpha(m)$ depending only on the dimension such that, for every $k \in \mathbb{N}^*$, there exists a family $\{(F_i, G_i)\}_{i=1}^k$ of capacitors with the following properties*:

(I) $\mu_\phi(F_i) > c(m)\dfrac{\mu_\phi(M)}{k}$,

(II) $\operatorname{cap}_\phi(F_i, G_i) \leq \dfrac{\mu_\phi(M)}{k}\left[\dfrac{1}{r_0^2}\left(\dfrac{V_\phi([g])}{\mu_\phi(M)}\right)^{2/m} + \alpha(m)\left(\dfrac{k}{\mu_\phi(M)}\right)^{2/m}\right]$, *where* $1/r_0 = 1600\max\{\sigma, 1\}$.

Proof. We consider the Bakry–Émery manifold (M, g, ϕ) as the metric measure space (M, d_{g_0}, μ_ϕ) where $g_0 \in [g]$ with $\mathrm{Ricci}_\phi(M, g_0) \geq -(m-1)$ and μ_ϕ is the weighted measure with respect to the metric g. According to Lemma 2.4, this space has the $(2, N, \xi)$-covering property with $\xi = \min\{1/\sigma, 1\}$. Having Proposition 2.1 and Lemma 2.4, and following steps analogous to those in Lemma 2.2 we see, for every $k \in \mathbb{N}^*$, there exists a family of k mutually disjoint capacitors $\{F_i, G_i\}$ satisfying the following properties:

- $\mu_\phi(F_i) \geq c(m)\mu_\phi(M)/k$, where $c(m) \in (0, 1)$ is a positive constant depending only on the dimension, and $\mu_\phi(G_i) \leq \mu_\phi(M)/k$. Either
  (a) all the F_i are annuli with outer radii smaller than ξ, $G_i = {}^2F_i$, and
  $$\mathrm{cap}_\phi^{(m)}(F_i, G_i) \leq C(m),$$
  where $C(m)$ is the constant defined in (20);

  or

  (b) all the F_i are domains in M, $G_i = F_i^{r_0}$ is the r_0-neighborhood of F_i, and $\mathrm{cap}_\phi^{(m)}(F_i, G_i) \leq r_0^{-2} V_\phi([g])/k$, with $r_0 = \xi/1600$.

Hence, $\mathrm{cap}_\phi^{(m)}(F_i, G_i) \leq C(m) + r_0^{-2} V_\phi([g])/k$. Now, for every $\varepsilon > 0$, we consider a family of functions $\{f_i\}_{i=1}^k$, $f_i \in \mathcal{T}(F_i, G_i)$ such that

$$\int_M |\nabla_g f_i|^m e^{-\phi}\, d\mu_g \leq \mathrm{cap}_\phi^{(m)}(F_i, G_i) + \varepsilon.$$

We repeat the same argument as before.

$$\begin{aligned}
\mathrm{cap}_\phi(F_i, G_i) &\leq \int_M |\nabla_g f_i|^2 e^{-\phi} d\mu_g \\
&\leq \left(\int_M |\nabla_g f_i|^m e^{-\phi} d\mu_g\right)^{2/m} \left(\int_M 1_{\operatorname{supp} f_i} e^{-\phi} d\mu_g\right)^{1-2/m} \\
&\leq \left[C(m)^{2/m} + \frac{1}{r_0^2}\left(\frac{V_\phi([g])}{k}\right)^{2/m} + \varepsilon^{2/m}\right]\left(\frac{\mu_\phi(M)}{k}\right)^{1-2/m}.
\end{aligned}$$

Having $1/r_0 = 1600/\xi = 1600 \max\{\sigma, 1\}$ and letting ε tend to zero, we obtain property (II). This completes the proof. □

3. Eigenvalues of Schrödinger operators

In this section, we prove Theorems 1.1 and 1.2. The idea of the proof is to construct a suitable family of test functions to be used in the variational characterization of the eigenvalues. Due to the min-max Theorem, we have the following variational characterization for the eigenvalues of the Schrödinger operator $L = \Delta_g + q$:

$$\lambda_k(\Delta_g + q) = \min_{V_k} \max_{0 \neq f \in V_k} \frac{\int_M |\nabla_g f|^2 \, d\mu_g + \int_M f^2 q \, d\mu_g}{\int_M f^2 \, d\mu_g},$$

where V_k is a k-dimensional linear subspace of $H^1(M)$ and μ_g is the Riemannian measure corresponding to the metric g.

According to this variational formula, for every family $\{f_i\}_{1=1}^{k}$ of disjointly supported test functions, one has

$$\lambda_k(\Delta_g + q) \leq \max_{i \in \{1,\dots,k\}} \frac{\int_M |\nabla_g f_i|^2 \, d\mu_g + \int_M f_i^2 q \, d\mu_g}{\int_M f_i^2 \, d\mu_g}. \tag{21}$$

The potential $q \in C^0(M)$ is a signed function (notice that we can assume $q \in L^1(M)$ as well). We define a signed measure σ associated to the potential q by

$$\sigma(A) = \int_A q \, d\mu_g \quad \text{for every measurable subset } A \text{ of } X.$$

For any signed measure ν we write $\nu = \nu^+ - \nu^-$, where ν^+ and ν^- are the positive and negative parts of ν, respectively. For any signed measure ν and $0 \leq \delta \leq 1$ we define a new signed measure ν_δ as $\nu_\delta := \delta\nu^+ - \nu^-$.

Let μ and ν be two signed measures on M. Then, according to [Grigor'yan et al. 2004, Lemma 4.3], we have

$$(\mu + \nu)_\delta \geq \mu_\delta + \nu_\delta. \tag{22}$$

Proof of Theorem 1.1. For a real number $\lambda \in \mathbb{R}$ define $\mu_\lambda := (\lambda\mu_g - \sigma)^+$ as a nonatomic Borel measure on M. We apply Lemma 2.2 to (M, g, μ_λ). Thus, for every $k \in \mathbb{N}^*$ and every $\lambda \in \mathbb{R}$, there exists a family $\{(F_i, G_i)\}_{i=1}^{2k}$ of $2k$ capacitors satisfying properties (I) and (II) of Lemma 2.2.

From now on, we take $\lambda := \lambda_k = \lambda_k(L)$. Property (I) yields

$$(\lambda_k\mu_g - \sigma)^+(F_i) \geq c(m)\frac{(\lambda_k\mu_g - \sigma)^+(M)}{2k}.$$

The measure $(\lambda_k\mu_g - \sigma)^-$ is also nonatomic. Since G_i are mutually disjoint, up to reordering, the first k of them satisfy

$$(\lambda_k\mu_g - \sigma)^-(G_i) \leq \frac{(\lambda_k\mu_g - \sigma)^-(M)}{k}, \quad i \in \{1, \dots, k\}.$$

Therefore

$$\begin{aligned}(\lambda_k\mu_g - \sigma)^-(G_i) &- (\lambda_k\mu_g - \sigma)^+(F_i) \\ &\leq \frac{(\lambda_k\mu_g - \sigma)^-(M)}{k} - c(m)\frac{(\lambda_k\mu_g - \sigma)^+(M)}{2k}.\end{aligned} \tag{23}$$

For every $\epsilon > 0$ and every $1 \le i \le k$, we choose $f_i \in \mathcal{T}(F_i, G_i)$ such that

$$\int_M |\nabla_g f_i|^2 \, d\mu_g \le \operatorname{cap}_g(F_i, G_i) + \epsilon. \tag{24}$$

Inequality (21) implies that there exists $i \in \{1, \dots, k\}$ so that

$$\lambda_k \int_M f_i^2 \, d\mu_g \le \int_M |\nabla_g f_i|^2 \, d\mu_g + \int_M f_i^2 q \, d\mu_g.$$

Hence, having Lemma 2.2 and inequality (23), we get

$$\begin{aligned}
0 &\le \int_M |\nabla_g f_i|^2 \, d\mu_g - \int_M f_i^2(\lambda_k - q) \, d\mu_g \\
&\le \operatorname{cap}_g(F_i, G_i) + \epsilon - \int_M f_i^2(\lambda_k - q) \, d\mu_g \\
&\le \frac{\mu_g(M)}{2k}\left[\frac{1}{r_0^2}\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + \alpha(m)\left(\frac{2k}{\mu_g(M)}\right)^{2/m}\right] + \epsilon \\
&\quad + \int_M f_i^2(\lambda_k - q)^- \, d\mu_g - \int_M f_i^2(\lambda_k - q)^+ \, d\mu_g \\
&\le \frac{\mu_g(M)}{2k}\left[\frac{1}{r_0^2}\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + \alpha(m)\left(\frac{2k}{\mu_g(M)}\right)^{2/m}\right] + \epsilon \\
&\quad + \frac{(\lambda_k \mu_g - \sigma)^-(M)}{k} - c(m)\frac{(\lambda_k \mu_g - \sigma)^+(M)}{2k}.
\end{aligned} \tag{25}$$

We now estimate the last two terms of this inequality considering two alternatives.

Case 1. If $\lambda_k = \lambda_k(L)$ is positive, then, applying (22) for the measure $\lambda_k \mu_g$ and signed measure $-\sigma$ with $\delta = c(m)/2$, we get

$$\begin{aligned}
&\frac{c(m)}{2}(\lambda_k \mu_g - \sigma)^+(M) - (\lambda_k \mu_g - \sigma)^-(M) \\
&\qquad \ge \frac{c(m)}{2}\sigma^-(M) - \sigma^+(M) + \frac{c(m)}{2}\lambda_k \mu_g(M).
\end{aligned} \tag{26}$$

Substituting (26) in (25) and letting ϵ tend to zero gives

$$\begin{aligned}
\lambda_k \le{}& \frac{(2/c(m))\sigma^+(M) - \sigma^-(M)}{\mu_g(M)} \\
&+ \frac{1}{c(m) r_0^2}\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + \frac{\alpha(m)}{c(m)}\left(\frac{2k}{\mu_g(M)}\right)^{2/m}.
\end{aligned} \tag{27}$$

Case 2. If $\lambda_k = \lambda_k(L)$ is nonpositive, applying (22) for the signed measures $\lambda_k \mu_g$ and $-\sigma$ with $\delta = c(m)/2$ implies

$$\frac{c(m)}{2}(\lambda_k \mu_g - \sigma)^+(M) - (\lambda_k \mu_g - \sigma)^-(M) \ge \frac{c(m)}{2}\sigma^-(M) - \sigma^+(M) + \lambda_k \mu_g(M).$$

Substituting this in (25) and letting ϵ go to zero gives

$$(28)\quad \lambda_k \le \frac{\sigma^+(M) - (c(m)/2)\sigma^-(M)}{\mu_g(M)} + \frac{1}{2r_0^2}\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + \frac{\alpha(m)}{2}\left(\frac{2k}{\mu_g(M)}\right)^{2/m}.$$

Therefore, $\lambda_k(L)$ is smaller than the sum of the right sides of inequalities (27) and (28). We finally obtain inequality (6) with, for example, $\alpha_m = c(m)/4$. □

Proof of Theorem 1.2. We partly follow the spirit of' the proof of [Grigor'yan et al. 2004, Theorem 5.15]. Take the metric measure space (M, g, μ_g). By Lemma 2.2, for every $k \in N^*$ there is a family of $2k$ disjoint capacitors $\{(F_i, G_i)\}_{i=1}^{2k}$ that satisfies properties (I) and (II) of Lemma 2.2. For every $\varepsilon > 0$, let $\{f_i\}_{i=1}^{2k}$ be a family of test functions with $2f_i \in \mathcal{T}(F_i, G_i)$ and $4\int_M |\nabla_g f_i|^2\, d\mu_g \le \operatorname{cap}_g(F_i, G_i) + \varepsilon$. We claim that this family satisfies the following property:

$$(29)\qquad \sum_{i=1}^{2k} \int_M f_i^2 q\, d\mu_g \le \sum_{i=1}^{2k} \int_M |\nabla_g f_i|^2\, d\mu_g + \int_M q\, d\mu_g.$$

If we have inequality (29),

$$(30)\qquad \begin{aligned}\sum_{i=1}^{2k} \int_M (|\nabla_g f_i|^2 + f_i^2 q)\, d\mu_g &\le 2\sum_{i=1}^{2k} \int_M |\nabla_g f_i|^2\, d\mu_g + \int_M q\, d\mu_g \\ &\le k \max_i \operatorname{cap}_g(F_i, G_i) + k\varepsilon + \int_M q\, d\mu_g.\end{aligned}$$

By the assumption, $\int_M (|\nabla_g f_i|^2 + f_i^2 q)\, d\mu_g$ is positive for each $1 \le i \le 2k$. Therefore, at least k of them (up to reordering we assume that it's the first k) satisfy the inequality

$$(31)\qquad \int_M (|\nabla_g f_i|^2 + f_i^2 q)\, d\mu_g \le \max_i \operatorname{cap}_g(F_i, G_i) + \varepsilon + \frac{\int_M q\, d\mu_g}{k}.$$

Inequality (31), together with the bounds of $\operatorname{cap}_g(F_i, G_i)$ and $\mu_g(F_i)$ given in Lemma 2.2 and properties (I) and (II), leads to

$$\begin{aligned}\lambda_k(L) &\le \max_i \frac{\int_M |\nabla_g f_i|^2 d\mu_g + \int_M f_i^2 q\, d\mu_g}{\int_M f_i^2 d\mu_g} \\ &\le \frac{\max_i \operatorname{cap}_g(F_i, G_i) + \varepsilon + (1/k)\int_M q\, d\mu_g}{\mu_g(F_i)} \\ &\le \frac{1}{c(m) r_0^2}\left(\frac{V([g])}{\mu_g(M)}\right)^{2/m} + \alpha(m)\left(\frac{2k}{\mu_g(M)}\right)^{2/m} + \frac{2k\varepsilon}{c(m)\mu_g(M)} + \frac{2\int_M q\, d\mu_g}{c(m)\mu_g(M)}.\end{aligned}$$

Hence we get the desired inequality as ε tends to zero. It remains to prove inequality (29) which is proved in [Grigor'yan et al. 2004, Section 5]; however, for the reader's convenience we repeat the proof. We define the function h by the identity

$$\sum_{i=1}^{2k} f_i^2 + h^2 = 1. \tag{32}$$

Since $f_1, \dots, f_{2k}$ are disjointly supported and $0 \le f_i \le 1/2$, $h \ge 1/2$. We now estimate the left side of inequality (29).

$$\begin{aligned}\int_M \Big(\sum_{i=1}^{2k} f_i^2 + h^2 - h^2\Big) q \, d\mu_g &= \int_M q \, d\mu_g - \int_M h^2 q \, d\mu_g \\ &\le \int_M q \, d\mu_g + \int_M |\nabla h|^2 \, d\mu_g,\end{aligned} \tag{33}$$

where the last inequality comes from the fact that the Schrödinger operator L is positive. Identity (32) implies

$$-2h\nabla_g h = -\nabla_g h^2 = \sum_{i=1}^{2k} \nabla_g f_i^2 = 2\sum_{i=1}^{2k} f_i \nabla_g f_i.$$

Therefore,

$$|\nabla_g h|^2 \le |2h\nabla_g h|^2 = \sum_{i=1}^{2k} |\nabla_g f_i^2|^2 = 4\sum_{i=1}^{2k} |f_i \nabla_g f_i|^2 \le \sum_{i=1}^{2k} |\nabla_g f_i|^2. \tag{34}$$

Combining inequalities (33) and (34) we get inequality (29). □

4. Eigenvalues of Bakry–Émery Laplace operators

In this section we consider eigenvalues of the Bakry–Émery Laplace operator Δ_ϕ on a Bakry–Émery manifold (M, g, ϕ), where M is a compact m-dimensional Riemannian manifold and $\phi \in C^2(M)$. We denote the weighted measure on M by μ_ϕ with

$$\mu_\phi(A) = \int_A e^{-\phi} \, d\mu_g \quad \text{for every Borel subset } A \text{ of } M.$$

Proof of Theorem 1.5. As we mentioned in the introduction, one can see that $\Delta_\phi = \Delta_g + \nabla_g \phi \cdot \nabla_g$ is unitarily equivalent to the positive Schrödinger operator $L = \Delta_g + \frac{1}{2}\Delta_g \phi + \frac{1}{4}|\nabla_g \phi|^2$. Therefore, Theorem 1.2 yields

$$\begin{aligned}\lambda_k(\Delta_\phi) \le A_m \frac{1}{\mu_g(M)} \int_M \big(\tfrac{1}{2}\Delta_g \phi + \tfrac{1}{4}|\nabla_g \phi|^2\big) \, d\mu_g \\ + B_m \Big(\frac{V([g])}{\mu_g(M)}\Big)^{2/m} + C_m \Big(\frac{k}{\mu_g(M)}\Big)^{2/m}.\end{aligned}$$

Now Stokes' theorem implies that $\int_M \Delta_g\phi \, d\mu_g = 0$. This gives the result. $\square$

For the proof of Theorem 1.7, we use the characteristic variational formula for the Bakry–Émery Laplacian; see for example [Lu and Rowlett 2012, Proposition 1; Rowlett 2010, Proposition 4].

$$\lambda_k(\Delta_\phi) = \inf_{V_k} \sup_{f \in V_k} \frac{\int_M |\nabla_g f|^2 e^{-\phi} \, d\mu_g}{\int_M f^2 e^{-\phi} \, d\mu_g}, \tag{35}$$

where V_k is a k-dimensional linear subspace of $H^1(M, \mu_\phi)$.

Proof of Theorem 1.7. According to Lemma 2.5, for $k \in \mathbb{N}^*$ we have a family of k capacitors satisfying properties (I) and (II). For every $\varepsilon > 0$, take $f_i \in \mathcal{T}(F_i, G_i)$, $1 \le i \le k$, so that

$$\int_M |\nabla_g f_i|^2 e^{-\phi} d\mu_g \le \operatorname{cap}_\phi(F_i, G_i) + \varepsilon.$$

Hence, the characteristic variational formula (35) gives

$$\lambda_k(\Delta_\phi) \le \max_i \frac{\int_M |\nabla_g f_i|^2 e^{-\phi} \, d\mu_g}{\int_M f_i^2 e^{-\phi} \, d\mu_g} \le \max_i \frac{\operatorname{cap}_\phi(F_i, G_i) + \varepsilon}{\mu_\phi(F_i)}.$$

Having the properties (I) and (II), we get

$$\lambda_k(\Delta_\phi) \le A(m) \max\{\sigma^2, 1\} \left(\frac{V_\phi([g])}{\mu_\phi(M)} \right)^{2/m} + B(m) \left(\frac{k}{\mu_\phi(M)} \right)^{2/m} + \frac{k\varepsilon}{c(m)\mu_\phi(M)}.$$

Letting ε go to zero, we get the desired inequality. $\square$

Appendix: Buser type upper bound on Bakry–Émery manifolds

Here, we present a direct and simple proof of a weaker version of Corollary 1.9. The idea behind this proof was used by Buser [1979, Satz 7], Cheng [1975], and Li and Yau [1980] in the case of the Laplace–Beltrami operator. It is based on constructing a family of balls as capacitors which will be the support of test functions. We can successfully apply this idea in the case of the Bakry–Émery Laplace operator.

Theorem A.1 (Buser type upper bound). *Let (M, g, ϕ) be a compact Bakry–Émery manifold with* $\operatorname{Ricci}_\phi(M) > -\kappa^2(m-1)$ *and $|\nabla_g \phi| \le \sigma$ for some $\kappa \ge 0$ and $\sigma \ge 0$. There are positive constants $A(m)$ and $B(m)$ such that, for every $k \in \mathbb{N}^*$,*

$$\lambda_k(\Delta_\phi) \le A(m) \max\{\sigma, \kappa\}^2 + B(m) \left(\frac{k}{\mu_\phi(M)} \right)^{2/m}.$$

To see that the above theorem is weaker than Corollary 1.9, consider the case where $\operatorname{Ricci}_\phi(M, g)$ is nonnegative. Indeed, the upper bound in Theorem A.1 still depends on σ while Corollary 1.9 provides an upper bound which depends only on the dimension.

Proof. Since $\mathrm{Ricci}_\phi(M) > -\kappa^2(m-1)$ and $|\nabla_g \phi| \le \sigma$, the comparison theorem gives us the following inequalities for every $0 < r \le \xi = \min\{1/\sigma, 1/\kappa\}$ (with $\xi = \infty$ if $\sigma = \kappa = 0$):

$$\frac{\mu_\phi(B(x,r))}{\mu_\phi(B(x,r/2))} \le e^{\sigma r} \frac{v(m,r,-\kappa^2)}{v(m,r/2,-\kappa^2)} \le \sup_{r\in(0,\xi)} e^{\sigma r} \frac{v(m,r,-\kappa^2)}{v(m,r/2,-\kappa^2)} =: c_1(m)$$

and

$$\mu_\phi(B(x,r)) \le e^{\sigma r} v(m,r,-\kappa^2) \le \sup_{s\in(0,\xi)} e^{\sigma s} v(m,s,-\kappa^2) r^m =: c_2(m) r^m. \tag{36}$$

Given $k \in \mathbb{N}^*$, let $\rho(k)$ be the positive number defined by

$$\rho(k) = \sup\{r : \text{there exist } p_1, \dots, p_k \in M \text{ with } d_g(p_i, p_j) > r \text{ for all } i \ne j\}.$$

We consider two cases.

Case 1. Let $\rho(k) \ge \xi$. For every $r < \xi$, there are k points $p_1, \dots, p_k$ with $B(p_i, r/2) \operatorname{cap} B(p_j, r/2) = \varnothing$ for all $i \ne j$. For each $i \in \{1, \dots, k\}$, we consider a plateau function $f_i \in \mathcal{T}(B(p_i, r/4), B(p_i, r/2))$, $1 \le i \le k$, defined as in (19). Then, for every $1 \le i \le k$ and every $r < \xi$,

$$\frac{\int_M |\nabla_g f_i|^2 e^{-\phi}\, d\mu_g}{\int_M f_i^2 e^{-\phi}\, d\mu_g} \le \frac{16}{r^2} \frac{\mu_\phi(B(p_i, r/2))}{\mu_\phi(B(p_i, r/4))} \le c_1(m) \frac{16}{r^2}.$$

Therefore, letting r tend to ξ, one has

$$\frac{\int_M |\nabla_g f_i|^2 e^{-\phi}\, d\mu_g}{\int_M f_i^2 e^{-\phi}\, d\mu_g} \le c_1(m) \frac{16}{\xi^2} \le A(m) \max\{\sigma, \kappa\}^2.$$

Case 2. Let $\rho(k) < \xi$. Take $r < \rho(k)$ very close to $\rho(k)$. As in Case 1, there are k points $p_1, \dots, p_k$ with $B(p_i, r/2) \operatorname{cap} B(p_j, r/2) = \varnothing$ for all $i \ne j$. Repeating the same argument, we get, for every $1 \le i \le k$,

$$\frac{\int_M |\nabla_g f_i|^2 e^{-\phi}\, d\mu_g}{\int_M f_i^2 e^{-\phi}\, d\mu_g} \le c_1(m) \frac{16}{r^2}.$$

Therefore, for every $1 \le i \le k$,

$$\frac{\int_M |\nabla_g f_i|^2 e^{-\phi}\, d\mu_g}{\int_M f_i^2 e^{-\phi}\, d\mu_g} \le c_1(m) \frac{16}{\rho(k)^2}.$$

We now estimate $\rho(k)$. Let $\rho(k) < s < \xi$ and n be the maximal number of points $q_1, \dots, q_n \in M$ so that $d(q_i, q_j) > s$ for all $i \ne j$. Of course $n \le k$ and because of the maximality of n, the balls $\{B(q_i, s)\}_{i=1}^n$ cover M. Hence, according to inequality (36),

$$\mu_\phi(M) \le \sum_{i=1}^{n} \mu_\phi(B(q_i, s)) \le nc_2(m)s^m \le kc_2(m)s^m.$$

Thus, letting s tend to $\rho(k)$, we get

$$\frac{1}{\rho(k)^2} \le c_2(m)^{2/m} \left(\frac{k}{\mu_\phi(M)}\right)^{2/m}.$$

Therefore,

$$\frac{\int_M |\nabla_g f_i|^2 e^{-\phi}\, d\mu_g}{\int_M f_i^2 e^{-\phi}\, d\mu_g} \le 16c_1(m)c_2(m)^{2/m} \left(\frac{k}{\mu_\phi(M)}\right)^{2/m}.$$

In conclusion, we obtain

$$\lambda_k(\Delta_\phi) \le \max_i \frac{\int_M |\nabla_g f_i|^2 e^{-\phi}\, d\mu_g}{\int_M f_i^2 e^{-\phi}\, d\mu_g} \le A(m) \max\{\sigma, \kappa\}^2 + B(m) \left(\frac{k}{\mu_\phi(M)}\right)^{2/m}. \quad \square$$

Acknowledgements

This paper is part of the author's PhD thesis under the direction of Professors Bruno Colbois (Neuchâtel University), Ahmad El Soufi (François Rabelais University), and Alireza Ranjbar-Motlagh (Sharif University of Technology). The author acknowledges their support and encouragement and thanks Bruno Colbois and Ahmad El Soufi for suggesting the problem and helpful discussions. She is also very grateful to the referee for helpful comments.

References

[Bakry and Émery 1985] D. Bakry and M. Émery, "Diffusions hypercontractives", pp. 177–206 in *Séminaire de probabilités, XIX* (Strasbourg, 1983/84), edited by J. Azéma and M. Yor, Lecture Notes in Math. **1123**, Springer, Berlin, 1985. MR 88j:60131 Zbl 0561.60080

[Bérard 1986] P. H. Bérard, *Spectral geometry: Direct and inverse problems*, Monografías de Matemática **41**, Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1986. MR 88d:58120 Zbl 0658.58003

[Buser 1979] P. Buser, "Beispiele für λ_1 auf kompakten Mannigfaltigkeiten", *Math. Z.* **165**:2 (1979), 107–133. MR 80d:58070 Zbl 0377.35058

[Cheng 1975] S. Y. Cheng, "Eigenvalue comparison theorems and its geometric applications", *Math. Z.* **143**:3 (1975), 289–297. MR 51 #14170 Zbl 0329.53035

[Colbois and Dodziuk 1994] B. Colbois and J. Dodziuk, "Riemannian metrics with large λ_1", *Proc. Amer. Math. Soc.* **122**:3 (1994), 905–906. MR 95a:58130 Zbl 0820.58056

[Colbois and Maerten 2008] B. Colbois and D. Maerten, "Eigenvalues estimate for the Neumann problem of a bounded domain", *J. Geom. Anal.* **18**:4 (2008), 1022–1032. MR 2009f:35243 Zbl 1158.58014

[El Soufi and Ilias 1992] A. El Soufi and S. Ilias, "Majoration de la seconde valeur propre d'un opérateur de Schrödinger sur une variété compacte et applications", *J. Funct. Anal.* **103**:2 (1992), 294–316. MR 93g:58150 Zbl 0766.58055

[Grigor'yan et al. 2004] A. Grigor'yan, Y. Netrusov, and S.-T. Yau, "Eigenvalues of elliptic operators and geometric applications", pp. 147–217 in *Surveys in differential geometry*, edited by A. Grigor'yan and S.-T. Yau, Surv. Differ. Geom. **9**, Int. Press, Somerville, MA, 2004. MR 2007f:58039 Zbl 1061.58027

[Hassannezhad 2011] A. Hassannezhad, "Conformal upper bounds for the eigenvalues of the Laplacian and Steklov problem", *J. Funct. Anal.* **261**:12 (2011), 3419–3436. MR 2012i:58024 Zbl 1232.58023

[Hassannezhad 2012] A. Hassannezhad, *Bornes supérieures pour les valeurs propres des opérateurs naturels sur des variétés riemannienes compactes*, Ph.D. thesis, Université de Neuchatel, 2012, Available at http://doc.rero.ch/record/30550/files/00002282.pdf. Text mostly in English.

[Korevaar 1993] N. Korevaar, "Upper bounds for eigenvalues of conformal metrics", *J. Differential Geom.* **37**:1 (1993), 73–93. MR 94d:58153 Zbl 0794.58045

[Li and Yau 1980] P. Li and S. T. Yau, "Estimates of eigenvalues of a compact Riemannian manifold", pp. 205–239 in *Geometry of the Laplace operator* (Honolulu, Hawaii, 1979), edited by R. Osserman and A. Weinstein, Proc. Sympos. Pure Math. **36**, Amer. Math. Soc., Providence, R.I., 1980. MR 81i:58050 Zbl 0441.58014

[Li and Yau 1982] P. Li and S. T. Yau, "A new conformal invariant and its applications to the Willmore conjecture and the first eigenvalue of compact surfaces", *Invent. Math.* **69**:2 (1982), 269–291. MR 84f:53049 Zbl 0503.53042

[Lott 2007] J. Lott, "Optimal transport and Ricci curvature for metric-measure spaces", pp. 229–257 in *Surveys in differential geometry*, edited by J. Cheeger and K. Grove, Surv. Differ. Geom. **11**, Int. Press, Somerville, MA, 2007. MR 2009i:53035 Zbl 1155.53026

[Lott and Villani 2009] J. Lott and C. Villani, "Ricci curvature for metric-measure spaces via optimal transport", *Ann. of Math.* (2) **169**:3 (2009), 903–991. MR 2010i:53068 Zbl 1178.53038

[Lu and Rowlett 2012] Z. Lu and J. Rowlett, "Eigenvalues of collapsing domains and drift Laplacians", *Math. Res. Lett.* **19**:3 (2012), 627–648.

[Munteanu and Wang 2012] O. Munteanu and J. Wang, "Analysis of weighted Laplacian and applications to Ricci solitons", *Comm. Anal. Geom.* **20**:1 (2012), 55–94. MR 2903101 Zbl 1245.53039

[Rowlett 2010] J. Rowlett, "La géométrie de Bakry–Émery et l'écart fondamental", pp. 147–157 in *Actes du Séminaire de Théorie Spectrale et Géometrie* (St. Martin d'Héres, 2009–2010), Sémin. Théor. Spectr. Géom. **28**, Institut Fourier, St. Martin d'Héres, 2010. MR 2012j:35289 Zbl 1223.58027

[Setti 1998] A. G. Setti, "Eigenvalue estimates for the weighted Laplacian on a Riemannian manifold", *Rend. Sem. Mat. Univ. Padova* **100** (1998), 27–55. MR 2000a:58082 Zbl 0922.58084

[Su and Zhang 2012] Y.-H. Su and H.-C. Zhang, "Rigidity of manifolds with Bakry–Émery Ricci curvature bounded below", *Geom. Dedicata* **160** (2012), 321–331. MR 2970058 Zbl 1257.53064

[Wei and Wylie 2009] G. Wei and W. Wylie, "Comparison geometry for the Bakry–Émery Ricci tensor", *J. Differential Geom.* **83**:2 (2009), 377–405. MR 2011a:53064 Zbl 1189.53036

[Wu 2010] J.-Y. Wu, "Upper bounds on the first eigenvalue for a diffusion operator via Bakry–Émery Ricci curvature", *J. Math. Anal. Appl.* **361**:1 (2010), 10–18. MR 2010m:58044 Zbl 1183.58028

[Wu 2012] J.-Y. Wu, "Upper bounds on the first eigenvalue for a diffusion operator via Bakry–Émery Ricci curvature II", preprint, 2012. arXiv 1010.4175v4

Received July 16, 2012. Revised December 26, 2012.

Asma Hassannezhad
Institut de Mathématiques
Université de Neuchâtel
Rue Emile-Argand 11, Case postale 158
CH-2009 Neuchâtel Switzerland
asma.hassannezhad@unine.ch

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.355

FOUR EQUIVALENT VERSIONS OF NONABELIAN GERBES

THOMAS NIKOLAUS AND KONRAD WALDORF

We recall and partially improve four versions of smooth, nonabelian gerbes: Čech cocycles, classifying maps, bundle gerbes, and principal 2-bundles. We prove that all four versions are equivalent, and so establish new relations between interesting recent developments. Prominent partial results that we prove are a bijection between the continuous and smooth nonabelian cohomology, and an explicit equivalence between bundle gerbes and principal 2-bundles as 2-stacks.

Thomas Nikolaus is supported by the Collaborative Research Centre 676 "Particles, Strings and the Early Universe — the Structure of Matter and Space-Time" and the cluster of excellence "Connecting particles with the cosmos".

MSC2010: primary 55R65; secondary 53C08, 55N05, 22A22.

Keywords: nonabelian gerbe, principal 2-bundle, 2-group, nonabelian cohomology, 2-stack.

1. Introduction

Let G be a Lie group and M be a smooth manifold. There are (among others) the following four ways to say what a *smooth G-bundle* over M is:

(1) *Čech 1-cocycles*: an open cover $\{U_i\}$ of M, and for each nonempty intersection $U_i \cap U_j$ a smooth map $g_{ij} : U_i \cap U_j \to G$ satisfying the cocycle condition

$$g_{ij} \cdot g_{jk} = g_{ik}.$$

(2) *Classifying maps*: a continuous map

$$f : M \to \mathfrak{B}G$$

to the classifying space $\mathfrak{B}G$ of the group G.

(3) *Bundle 0-gerbes*: a surjective submersion $\pi : Y \to M$ and a smooth map $g : Y \times_M Y \to G$ satisfying

$$\pi_{12}^* g \cdot \pi_{23}^* g = \pi_{13}^* g,$$

where $\pi_{ij} : Y \times_M Y \times_M Y \to Y \times_M Y$ denotes the projection to the i-th and j-th factors.

(4) *Principal bundles*: a surjective submersion $\pi : P \to M$ with a smooth action of G on P that preserves π, such that the map

$$P \times G \to P \times_M P : (p, g) \mapsto (p, p.g)$$

is a diffeomorphism.

It is well-known that these four versions of smooth G-bundles are all equivalent. Indeed, (1) forms the smooth G-valued Čech cohomology in degree one, whereas (2) is known to be equivalent to continuous G-valued Čech cohomology, which in turn coincides with smooth G-valued Čech cohomology. Further, (3) and (4) form equivalent categories, and isomorphism classes of the objects (3) are in bijection with equivalence classes of the cocycles (1).

In this article we provide an analogous picture for *smooth Γ-gerbes*, where Γ is a strict Lie 2-group. In particular, Γ can be the automorphism 2-group of an ordinary Lie group G, in which case the term "nonabelian G-gerbe" is commonly used. We compare the following four versions:

Version I: *Smooth, nonabelian Čech Γ-cocycles* (Definition 3.6). These form the classical, smooth groupoid-valued cohomology $\check{\mathrm{H}}^1(M, \Gamma)$ in the sense of Giraud [1971] and Breen [1990, Chapter 4; 1994].

Version II: *Classifying maps* (Definition 4.4). These are continuous maps $f : M \to \mathfrak{B}|\Gamma|$ to the classifying space of the geometric realization of Γ; such maps have been introduced and studied in [Baez and Stevenson 2009].

Version III: Γ-*bundle gerbes* (Definition 5.1.1). These have been developed by Aschieri, Cantini and Jurčo [Aschieri et al. 2005] as a generalization of the abelian bundle gerbes of Murray [1996]. Here we present an equivalent definition by applying a higher categorical version [Nikolaus and Schweigert 2011] of Grothendieck's stackification construction to the monoidal pre-2-stack of principal Γ-bundles.

Version IV: *Principal* Γ-2-*bundles* (Definition 6.1.5). These have been introduced in [Bartels 2006]; their total spaces are Lie groupoids on which the Lie 2-group Γ acts in a certain way. Compared to Bartels' definition, ours uses a stricter and easier notion of such an action.

We prove that all four versions are equivalent, and follow the same line of arguments as in the case of G-bundles outlined above:

- Baez and Stevenson have shown that homotopy classes of classifying maps of Version II are in bijection with the continuous groupoid-valued Čech cohomology $\check{\mathrm{H}}^1_c(M, \Gamma)$. We prove (Proposition 4.1) that the inclusion of *smooth* into *continuous* Čech Γ-cocycles induces a bijection

$$\check{\mathrm{H}}^1_c(M, \Gamma) \cong \check{\mathrm{H}}^1(M, \Gamma).$$

  These two results establish the equivalence between our Versions I and II (Theorem 4.6).

- Γ-bundle gerbes and principal Γ-2-bundles over M form bicategories. We prove (Theorem 7.0.1) that these bicategories are equivalent, and so establish the equivalence between Versions III and IV. Our proof provides explicit 2-functors in both directions.

- We prove the equivalence between Versions I and III by showing that nonabelian Γ-bundle gerbes are classified by the nonabelian cohomology group $\check{\mathrm{H}}^1(M, \Gamma)$ (Theorem 5.3.2).

The first aim of this paper is to simplify and clarify the notion of a nonabelian gerbe. This concerns the notion of a Γ-bundle gerbe (Version III), for which we give a new, conceptually clear, and manifestly 2-categorical definition. It also concerns the notion of a principal 2-bundle (Version IV), for which we provide a new definition that is carefully balanced between generality and simplicity.

The second aim of this paper is to make it possible to compare and transfer available results between the various versions. Indeed, none of the three equivalences above is available in the existing literature. As an example why such equivalences can be useful, we use Theorem 7.0.1 — the equivalence between Γ-bundle gerbes and principal Γ-2-bundles — in order to carry two facts about Γ-bundle gerbes over to principal Γ-2-bundles. We prove:

(1) Principal Γ-2-bundles form a 2-stack over smooth manifolds (Theorem 6.2.1). This is a new and evidently important result, since it explains precisely in which way one can *glue* 2-bundles from local patches.

(2) If Γ and Ω are weakly equivalent Lie 2-groups, the 2-stacks of principal Γ-2-bundles and principal Ω-2-bundles are equivalent (Theorem 6.2.3). This is another new result that generalizes the well-known fact that principal G-bundles and principal H-bundles form equivalent stacks, whenever G and H are isomorphic Lie groups.

The two facts about Γ-bundle gerbes (Theorems 5.1.5 and 5.2.2) on which these results are based are proved in an outmost abstract way: the first is a mere consequence of the definition of Γ-bundle gerbes that we give, namely via a 2-stackification procedure for principal Γ-bundles. The second follows from the fact that principal Γ-bundles and principal Ω-bundles form equivalent monoidal pre-2-stacks, which we deduce as a corollary of their description by anafunctors.

The present paper is part of a larger program. In a forthcoming paper, we address the discussion of nonabelian lifting problems, in particular string structures. In a second forthcoming paper we will present the picture of four equivalent versions in a setting *with connections*, based on the results of the present paper. Our motivation is to understand the role of 2-bundles with connections in higher gauge theories, where they serve as “B-fields”. Here, two (nonabelian) 2-groups are especially important, namely the string group [Baez et al. 2007] and the Jandl group [Nikolaus and Schweigert 2011]. More precisely, string-2-bundles appear in supersymmetric sigma models that describe fermionic string theories [Bunke 2011], while Jandl-2-bundles appear in unoriented sigma models that describe, e.g., bosonic type-I string theories [Schreiber et al. 2007].

This paper is organized as follows. In Section 2 we recall and summarize the theory of principal groupoid bundles and their description by anafunctors. The rest of the paper is based on this theory. In Sections 3–6 we introduce our four versions of smooth Γ-gerbes, and establish all but one equivalence. The remaining equivalence, the one between bundle gerbes and principal 2-bundles, is discussed in Section 7.

2. Preliminaries

There is no claim of originality in this section. Our sources are [Lerman 2010; Metzler 2003; Heinloth 2005; Moerdijk and Mrčun 2003]. A slightly different but equivalent approach is developed in [Murray et al. 2012].

2.1. ***Lie groupoids and groupoid actions on manifolds.*** We assume that the reader is familiar with the notions of Lie groupoids, smooth functors and smooth natural transformations. In this paper, the following examples of Lie groupoids appear:

Example 2.1.1. (a) Every smooth manifold M defines a Lie groupoid, denoted by M_{dis}, whose objects and morphisms are M, and all of whose structure maps are identities.

(b) Every Lie group G defines a Lie groupoid denoted by $\mathcal{B}G$, with one object, with G as its smooth manifold of morphisms, and with the composition $g_2 \circ g_1 := g_2 g_1$.

(c) Suppose X is a smooth manifold and $\rho : H \times X \to X$ is a smooth left action of a Lie group H on X. Then, a Lie groupoid $X/\!/H$ is defined with objects X and morphisms $H \times X$, and with

$$s(h, x) := x, \quad t(h, x) := \rho(h, x) \quad \text{and} \quad \mathrm{id}_x := (1, x).$$

The composition is

$$(h_2, x_2) \circ (h_1, x_1) := (h_2 h_1, x_1),$$

where $x_2 = \rho(h_1, x_1)$. The Lie groupoid $X/\!/H$ is called the *action groupoid* of the action of H on X.

(d) Let $t : H \to G$ be a homomorphism of Lie groups. Then,

$$\rho : H \times G \to G : (h, g) \mapsto (t(h)g)$$

defines a smooth left action of H on G. Thus, we have a Lie groupoid $G/\!/H$.

(e) To every Lie groupoid Γ one can associate an *opposite Lie groupoid* Γ^{op} which has the source and the target map exchanged.

We say that a *right action* of a Lie groupoid Γ on a smooth manifold M is a pair (α, ρ) consisting of smooth maps $\alpha : M \to \Gamma_0$ and $\rho : M \;{}_\alpha\!\times_t \Gamma_1 \to M$ such that

$$\rho(\rho(x, g), h) = \rho(x, g \circ h), \quad \rho(x, \mathrm{id}_{\alpha(x)}) = x \quad \text{and} \quad \alpha(\rho(x, g)) = s(g)$$

for all possible $g, h \in \Gamma_1$, $p \in \Gamma_0$ and $x \in M$. The map α is called *anchor*. Later on we will replace the letter ρ for the action by the symbol $\circ$ that denotes the composition of the groupoid. A *left action* of Γ on M is a right action of the opposite Lie groupoid Γ^{op}. A smooth map $f : M \to M'$ between Γ-spaces with actions (α, ρ) and (α', ρ') is called Γ-*equivariant* if

$$\alpha' \circ f = \alpha \quad \text{and} \quad f(\rho(x, g)) = \rho'(f(x), g).$$

Example 2.1.2. (1) Let Γ be a Lie groupoid. Then, Γ acts on the right on its morphisms Γ_1 by $\alpha := s$ and $\rho := \circ$. It acts on the left on its morphisms by $\alpha := t$ and $\rho := \circ$.

(2) Let G be a Lie group. Then, a right/left action of the Lie groupoid $\mathcal{B}G$ (see Example 2.1.1(b)) on M is the same as an ordinary smooth right/left action of G on M.

(3) Let X be a smooth manifold. A right/left action of X_{dis} (see Example 2.1.1(a)) on M is the same as a smooth map $\alpha : M \to X$.

2.2. ***Principal groupoid bundles.*** We give the definition of a principal bundle in exactly the same way as we are going to define principal 2-bundles in Section 6.

Definition 2.2.1. Let M be a smooth manifold, and let Γ be a Lie groupoid.

(1) A *principal* Γ*-bundle over* M is a smooth manifold P with a surjective submersion $\pi : P \to M$ and a right Γ-action (α, ρ) that respects the projection π, such that
$$\tau : P \,{}_{\alpha}\!\times_{t} \Gamma_1 \to P \times_M P : (p, g) \mapsto (p, \rho(p, g))$$
is a diffeomorphism.

(2) Let P_1 and P_2 be principal Γ-bundles over M. A *morphism* $\varphi : P_1 \to P_2$ is a Γ-equivariant smooth map that respects the projections to M.

Principal Γ-bundles over M form a category $\mathcal{B}\mathrm{un}_\Gamma(M)$. In fact, this category is a groupoid; i.e., all morphisms between principal Γ-bundles are invertible. There is an evident notion of a pullback f^*P of a principal Γ-bundle P over M along a smooth map $f : X \to M$, and similarly, morphisms between principal Γ-bundles pull back. These define a functor
$$f^* : \mathcal{B}\mathrm{un}_\Gamma(M) \to \mathcal{B}\mathrm{un}_\Gamma(X).$$
These functors make principal Γ-bundles a prestack over smooth manifolds. One can easily show that this prestack is a stack (for the Grothendieck topology of surjective submersions).

Example 2.2.2 (ordinary principal bundles). For G a Lie group, we have an equality of categories
$$\mathcal{B}\mathrm{un}_{\mathcal{B}G}(M) = \mathcal{B}\mathrm{un}_G(M);$$
i.e., Definition 2.2.1 reduces consistently to the definition of an ordinary principal G-bundle.

Example 2.2.3 (trivial principal groupoid bundles). For M a smooth manifold and $f : M \to \Gamma_0$ a smooth map, $P := M \,{}_f\!\times_t \Gamma_1$ and $\pi(m, g) := m$ define a surjective submersion, and $\alpha(m, g) := s(g)$ and $\rho((m, g), h) := (m, g \circ h)$ define a right action of Γ on P that preserves the fibers. The map τ we have to look at has the inverse
$$\tau^{-1} : P \times_M P \to P \,{}_{\pi}\!\times_t \Gamma_1 : ((m, g_1), (m, g_2)) \mapsto ((m, g_1), g_1^{-1} \circ g_2),$$

which is smooth. Thus we have defined a principal Γ-bundle, which we denote by $\mathbf{I}_f$ and which we call the *trivial bundle for the map* f. Any bundle that is isomorphic to a trivial bundle is called *trivializable*.

Example 2.2.4 (discrete structure groupoids). For X a smooth manifold, we have an equivalence of categories

$$\mathcal{B}\mathrm{un}_{X_{\mathrm{dis}}}(M) \cong C^\infty(M, X)_{\mathrm{dis}}.$$

Indeed, for a given principal X_{dis}-bundle P one observes that the anchor $\alpha : P \to X$ descends along the bundle projection to a smooth map $f : M \to X$, and that isomorphic bundles determine the same map. Conversely, one associates to a smooth map $f : M \to X$ the trivial principal X_{dis}-bundle $\mathbf{I}_f$ over M.

Example 2.2.5 (exact sequences). Let

$$1 \longrightarrow H \xrightarrow{t} G \xrightarrow{p} K \longrightarrow 1 \tag{2.2-1}$$

be an exact sequence of Lie groups, and let $\Gamma := G/\!\!/H$ be the action groupoid associated to the Lie group homomorphism $t : H \to G$ as explained in Example 2.1.1(d). In this situation, $p : G \to K$ is a surjective submersion, and

$$\alpha : G \to \Gamma_0 : g \mapsto g \quad \text{and} \quad \rho : G\,{}_{\alpha}\!\times_t \Gamma_1 \to G : (g, (h, g')) \mapsto g'$$

define a smooth right action of Γ on G that preserves p. The inverse of the map τ is

$$\tau^{-1} : G \times_K G \to G\,{}_{\alpha}\!\times_t \Gamma_1 : (g_1, g_2) \mapsto \big(g_1, (t^{-1}(g_1 g_2^{-1}), g_2)\big),$$

which is smooth because t is an embedding. Thus, G is a principal Γ-bundle over K.

Next we provide some elementary statements about trivial principal Γ-bundles.

Lemma 2.2.6. *A principal* Γ*-bundle over* M *is trivializable if and only if it has a smooth section.*

Proof. A trivial bundle $\mathbf{I}_f$ has the section

$$s_f : M \to \mathbf{I}_f : x \mapsto (x, \mathrm{id}_{f(x)}),$$

and so any trivializable bundle has a section. Conversely, suppose a principal Γ-bundle P has a smooth section $s : M \to P$. Then, with $f := \alpha \circ s$,

$$\varphi : \mathbf{I}_f \to P : (m, g) \mapsto \rho(s(m), g)$$

is an isomorphism. □

The following consequence shows that principal Γ-bundles of Definition 2.2.1 are locally trivializable in the usual sense.

Corollary 2.2.7. *Let P be a principal Γ-bundle over M. Then, every point $x \in M$ has an open neighborhood U over which P has a trivialization: a smooth map $f : U \to \Gamma_0$ and a morphism $\varphi : \mathbf{I}_f \to P|_U$.*

Proof. One can choose U such that the surjective submersion $\pi : U \to P$ has a smooth section. Then, Lemma 2.2.6 applies to the restriction $P|_U$. □

We determine the Hom-set $\mathcal{H}\mathrm{om}(\mathbf{I}_{f_1}, \mathbf{I}_{f_2})$ between trivial principal Γ-bundles defined by smooth maps $f_1, f_2 : M \to \Gamma_0$. To a bundle morphism $\varphi : \mathbf{I}_{f_1} \to \mathbf{I}_{f_2}$ one associates the smooth function $g : M \to \Gamma_1$ which is uniquely defined by the condition

$$(\varphi \circ s_{f_1})(x) = s_{f_2}(x) \circ g(x)$$

for all $x \in M$. It is straightforward to see that:

Lemma 2.2.8. *The above construction defines a bijection*

$$\mathcal{H}\mathrm{om}(\mathbf{I}_{f_1}, \mathbf{I}_{f_2}) \to \{g \in C^\infty(M, \Gamma_1) \mid s \circ g = f_1 \text{ and } t \circ g = f_2\},$$

under which identity morphisms correspond to constant maps and the composition of bundle morphisms corresponds to the pointwise composition of functions.

Finally, we consider the case of principal bundles for action groupoids.

Lemma 2.2.9. *Let $X /\!\!/ H$ be a smooth action groupoid. The category $\mathcal{B}\mathrm{un}_{X /\!\!/ H}(M)$ is equivalent to a category with:*

- *Objects: principal H-bundles P_H over M together with a smooth, H-antiequivariant map $f : P_H \to X$; i.e., $f(p \cdot h) = h^{-1} f(p)$.*
- *Morphisms: bundle morphisms $\varphi_H : P_H \to P'_H$ that respect the maps f and f'.*

Proof. For a principal $X /\!\!/ H$-bundle (P, α, ρ) we set $P_H := P$ with the given projection to M. The action of H on P_H is defined by

$$p \star h := \rho\big(p, (h, h^{-1} \cdot \alpha(p))\big).$$

This action is smooth, and it follows from the axioms of the principal bundle P that it is principal. The map $f : P_H \to X$ is the anchor α. The remaining steps are straightforward and left as an exercise. □

2.3. *Anafunctors.* An anafunctor is a generalization of a smooth functor between Lie groupoids, similar to a Morita equivalence, and also known as a Hilsum–Skandalis morphism. The idea goes back to [Bénabou 1973]; also see [Johnstone 1977]. The references for the following definitions are [Lerman 2010; Metzler 2003].

Definition 2.3.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be Lie groupoids.

(1) An *anafunctor* $F : \mathcal{X} \to \mathcal{Y}$ is a smooth manifold F, a left action (α_l, ρ_l) of $\mathcal{X}$ on F, and a right action (α_r, ρ_r) of $\mathcal{Y}$ on F such that the actions commute and $\alpha_l : F \to \mathcal{X}_0$ is a principal $\mathcal{Y}$-bundle over $\mathcal{X}_0$.

(2) A *transformation* between anafunctors $f : F \Rightarrow F'$ is a smooth map $f : F \to F'$ which is $\mathcal{X}$-equivariant, $\mathcal{Y}$-equivariant, and satisfies $\alpha'_l \circ f = \alpha_l$ and $\alpha'_r \circ f = \alpha_r$.

The smooth manifold F of an anafunctor is called its *total space*. Notice that the condition that the two actions on F commute implies that each respects the anchor of the other. For fixed Lie groupoids $\mathcal{X}$ and $\mathcal{Y}$, anafunctors $F : \mathcal{X} \to \mathcal{Y}$ and transformations form a category $\mathcal{A}\mathrm{na}^\infty(\mathcal{X}, \mathcal{Y})$. Since transformations are in particular morphisms between principal $\mathcal{Y}$-bundles, every transformation is invertible so that $\mathcal{A}\mathrm{na}^\infty(\mathcal{X}, \mathcal{Y})$ is in fact a groupoid.

Example 2.3.2 (anafunctors from ordinary functors). Given a smooth functor $\phi : \mathcal{X} \to \mathcal{Y}$, we obtain an anafunctor in the following way. We set $F := \mathcal{X}_0 \, {}_\phi\!\times_t \mathcal{Y}_1$ with anchors $\alpha_l : F \to \mathcal{X}_0$ and $\alpha_r : F \to \mathcal{Y}_0$ defined by $\alpha_l(x, g) := x$ and $\alpha_r(x, g) := s(g)$, and actions

$$\rho_l : \mathcal{X}_1 \, {}_s\!\times_{\alpha_l} F \to F \quad \text{and} \quad \rho_r : F \, {}_{\alpha_r}\!\times_t \mathcal{Y}_1 \to F$$

defined by $\rho_l(f, (x, g)) := (t(f), \phi(f) \circ g)$ and $\rho_r((x, g), f) := (x, g \circ f)$. In the same way, a smooth natural transformation $\eta : \phi \Rightarrow \phi'$ defines a transformation $f_\eta : F \Rightarrow F'$ by $f_\eta(x, g) := (x, \eta(x) \circ g)$. Conversely, one can show that an anafunctor comes from a smooth functor, if its principal Γ-bundle has a smooth section.

Example 2.3.3 (anafunctors with discrete source). For M a smooth manifold and Γ a Lie groupoid, we have an equality of categories

$$\mathcal{B}\mathrm{un}_\Gamma(M) = \mathcal{A}\mathrm{na}^\infty(M_{\mathrm{dis}}, \Gamma).$$

Further, *trivial* principal Γ-bundles correspond to smooth *functors*. In particular, with Example 2.2.2 we have:

(a) For G a Lie group and M a smooth manifold, an anafunctor $F : M_{\mathrm{dis}} \to \mathcal{B}G$ is the same as an ordinary principal G-bundle over M.

(b) For M and X smooth manifolds, an anafunctor $F : M_{\mathrm{dis}} \to X_{\mathrm{dis}}$ is the same as a smooth map.

Example 2.3.4 (anafunctors with discrete target). For Γ a Lie groupoid and M a smooth manifold, we have an equivalence of categories

$$C^\infty(\Gamma_0, M)^\Gamma_{\mathrm{dis}} \cong \mathcal{A}\mathrm{na}^\infty(\Gamma, M_{\mathrm{dis}})$$

where $C^\infty(\Gamma_0, M)^\Gamma$ denotes the set of smooth maps $f : \Gamma_0 \to M$ such that $f \circ s = f \circ t$ as maps $\Gamma_1 \to M$. The equivalence is induced by regarding a map $f \in C^\infty(\Gamma_0, M)^\Gamma$

as a smooth functor $f : \Gamma \to M_{\mathrm{dis}}$, which in turn induces an anafunctor. Conversely, an anafunctor $F : \Gamma \to M_{\mathrm{dis}}$ is in particular an M_{dis}-bundle over Γ_0, which is nothing but a smooth function $f : \Gamma_0 \to M$ by Example 2.2.4. The additional Γ-action assures the Γ-invariance of f.

Example 2.3.5 (anafunctors between one-object Lie groupoids). Let G and H be Lie groups, and let $\mathcal{B}G$ and $\mathcal{B}H$ be the associated one-object Lie groupoids (Example 2.1.1(b)). Then, there is an equivalence of categories

$$\mathrm{Hom}(G, H)/\!\!/H \cong \mathcal{A}\mathrm{na}^{\infty}(\mathcal{B}G, \mathcal{B}H),$$

where the action of H on $\mathrm{Hom}(G, H)$ is by pointwise conjugation. The functor which establishes this equivalence sends a smooth group homomorphism $\alpha : G \to H$ to the evident smooth functor $F_\alpha : \mathcal{B}G \to \mathcal{B}H$ and converts this into an anafunctor (Example 2.3.2). A morphism $h : \alpha_1 \to \alpha_2$ is sent to the natural transformation $\eta_h : F_{\alpha_1} \to F_{\alpha_2}$ whose component at the single object is the morphism $h \in H$. In order to see that this is essentially surjective, it suffices to notice that the principal H-bundle of any smooth anafunctor $F : \mathcal{B}G \to \mathcal{B}H$ has a section. The proof that the functor is full and faithful is straightforward.

For the following definition, we suppose $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are Lie groupoids, and $F : \mathcal{X} \to \mathcal{Y}$ and $G : \mathcal{Y} \to \mathcal{Z}$ are anafunctors given by $F = (F, \alpha_l, \rho_l, \alpha_r, \rho_r)$ and $G = (G, \beta_l, \tau_l, \beta_r, \tau_r)$.

Definition 2.3.6. The *composition* $G \circ F : \mathcal{X} \to \mathcal{Z}$ is the anafunctor defined in the following way:

(1) Its total space is

$$E := (F \,{}_{\alpha_r}\!\times_{\beta_l} G) / \sim$$

where $(f, \tau_l(h, g)) \sim (\rho_r(f, h), g)$ for all $h \in \mathcal{Y}_1$ with $\alpha_r(f) = t(h)$ and $\beta_l(g) = s(h)$.

(2) The anchors are $(f, g) \mapsto \alpha_l(f)$ and $(f, g) \mapsto \beta_r(g)$.

(3) The actions $\mathcal{X}_1 \,{}_s\!\times_\alpha E \to E$ and $E \,{}_\beta\!\times_t \mathcal{Z}_1 \to E$ are given, respectively, by

$$(\gamma, (f, g)) \mapsto (\rho_l(\gamma, f), g) \quad \text{and} \quad ((f, g), \gamma) \mapsto (f, \tau_r(g, \gamma)).$$

Remark 2.3.7. Lie groupoids, anafunctors and transformations form a bicategory. This bicategory is equivalent to the bicategory of differentiable stacks (also known as geometric stacks) [Pronk 1996].

In this article, anafunctors serve two purposes. The first is that one can use conveniently the composition of anafunctors to define *extensions* of principal groupoid bundles:

Definition 2.3.8. If $P : M_{\mathrm{dis}} \to \Gamma$ is a principal Γ-bundle over M, and $\Lambda : \Gamma \to \Omega$ is an anafunctor, then the principal Ω-bundle

$$\Lambda P := \Lambda \circ P : M_{\mathrm{dis}} \to \Omega$$

is called the *extension of P along Λ*.

Unwinding this definition, the principal Ω-bundle ΛP has the total space

$$\Lambda P = (P \,{}_{\alpha}\!\times_{\alpha_l} \Lambda) \,/\sim \tag{2.3-1}$$

where $(p, \rho_l(\gamma, \lambda)) \sim (\rho(p, \gamma), \lambda)$ for all $p \in P$, $\lambda \in \Lambda$ and $\gamma \in \Gamma_1$ with $\alpha(p) = t(\gamma)$ and $\alpha_l(\lambda) = s(\gamma)$. Here α is the anchor and ρ is the action of P, and $\Lambda = (\Lambda, \alpha_l, \alpha_r, \rho_l, \rho_r)$. The bundle projection is $(p, \lambda) \mapsto \pi(p)$, where π is the bundle projection of P, the anchor is $(p, \lambda) \mapsto \alpha_r(\lambda)$, and the action is

$$(p, \lambda) \circ \omega = (p, \rho_r(\lambda, \omega)).$$

Extensions of bundles are accompanied by extensions of bundle morphisms. If $\varphi : P_1 \to P_2$ is a morphism between Γ-bundles, a morphism $\Lambda\varphi : \Lambda P_1 \to \Lambda P_2$ is defined by $\Lambda\varphi(p_1, \lambda) := (\varphi(p_1), \lambda)$ in terms of (2.3-1). Summarizing, we have:

Lemma 2.3.9. *Let M be a smooth manifold and $\Lambda : \Gamma \to \Omega$ be an anafunctor. Then, extension along Λ is a functor*

$$\Lambda : \mathcal{B}\mathrm{un}_{\Gamma}(M) \to \mathcal{B}\mathrm{un}_{\Omega}(M).$$

Moreover, it commutes with pullbacks and so extends to a morphism between stacks.

Next we suppose that $t : H \to G$ is a Lie group homomorphism, and $G/\!\!/H$ is the associated action groupoid of Example 2.1.1(d). We look at the functor $\Theta : G/\!\!/H \to \mathcal{B}H$ which is defined by $\Theta(h, g) := h$. Combining Lemma 2.2.9 with the extension along Θ, we obtain:

Lemma 2.3.10. *The category $\mathcal{B}\mathrm{un}_{G/\!\!/H}(M)$ of principal $G/\!\!/H$-bundles over a smooth manifold M is equivalent to a category with*:

- *Objects*: *principal H-bundles P_H over M together with a section of $\Theta(P_H)$.*
- *Morphisms*: *morphisms φ of H-bundles so that $\Theta(\varphi)$ preserves the sections.*

The second motivation for introducing anafunctors is that they provide the inverses to certain smooth functors which are not necessarily equivalences of categories.

Definition 2.3.11. A smooth functor or anafunctor $F : \mathcal{X} \to \mathcal{Y}$ is called a *weak equivalence*, if there exists an anafunctor $G : \mathcal{Y} \to \mathcal{X}$ together with transformations $G \circ F \cong \mathrm{id}_{\mathcal{X}}$ and $F \circ G \cong \mathrm{id}_{\mathcal{Y}}$.

We have the following immediate consequence for the stack morphisms of Lemma 2.3.9.

Corollary 2.3.12. *Let $\Lambda : \Gamma \to \Omega$ be a weak equivalence between Lie groupoids. Then, extension of principal bundles along Λ is an equivalence $\Lambda : \mathcal{B}\mathrm{un}_\Gamma(M) \to \mathcal{B}\mathrm{un}_\Omega(M)$ of categories. Moreover, these define an equivalence between the stacks $\mathcal{B}\mathrm{un}_\Gamma(-)$ and $\mathcal{B}\mathrm{un}_\Omega(-)$.*

Concerning the claimed generalization of invertibility, we have the following well-known theorem; see [Lerman 2010, Lemma 3.34; Metzler 2003, Proposition 60].

Theorem 2.3.13. *A smooth functor $F : \mathcal{X} \to \mathcal{Y}$ is a weak equivalence if and only if the following two conditions are satisfied*:

(a) *It is smoothly essentially surjective*: *the map*

$$s \circ \mathrm{pr}_2 : \mathcal{X}_0 \, {}_{F_0}\!\times_t \mathcal{Y}_1 \to \mathcal{Y}_0$$

is a surjective submersion.

(b) *It is smoothly fully faithful*: *the diagram*

$$\begin{tikzcd} \mathcal{X}_1 \arrow[r, "F"] \arrow[d, "s\times t"'] & \mathcal{Y}_1 \arrow[d, "s\times t"] \\ \mathcal{X}_0 \times \mathcal{X}_0 \arrow[r, "F\times F"'] & \mathcal{Y}_0 \times \mathcal{Y}_0 \end{tikzcd}$$

is a pullback diagram.

Remark 2.3.14. One can show that any smooth functor $F : \mathcal{X} \to \mathcal{Y}$ that is a weak equivalence actually has a *canonical* inverse anafunctor.

2.4. ***Lie 2-groups and crossed modules.*** A (strict) *Lie 2-group* is a Lie groupoid Γ whose objects and morphisms are Lie groups, and all of whose structure maps are Lie group homomorphisms. One can conveniently bundle the multiplications and the inversions into smooth functors

$$m : \Gamma \times \Gamma \to \Gamma \quad \text{and} \quad i : \Gamma \to \Gamma.$$

Example 2.4.1. For A an abelian Lie group, we have that the Lie groupoid $\mathcal{B}A$ from Example 2.1.1(b) is a Lie 2-group. The condition that A is abelian is necessary.

Example 2.4.2. Let $t : H \to G$ be a homomorphism of Lie groups, and let $G /\!/ H$ be the corresponding Lie groupoid from Example 2.1.1(d). This Lie groupoid becomes a Lie 2-group if the following structure is given: a smooth left action of G on H by Lie group homomorphisms, denoted by $(g, h) \mapsto {}^g h$, satisfying

$$t({}^g h) = g t(h) g^{-1} \quad \text{and} \quad {}^{t(h)}x = h x h^{-1}$$

for all $g \in G$ and $h, x \in H$. Indeed, the objects G of $G/\!/H$ already form a Lie group, and the multiplication on the morphisms $H \times G$ of $G/\!/H$ is the semidirect product

$$(h_2, g_2) \cdot (h_1, g_1) = (h_2 {}^{g_2}h_1, g_2 g_1). \tag{2.4-1}$$

The homomorphism $t : H \to G$ together with the action of G on H is called a *smooth crossed module*. Summarizing, every smooth crossed module defines a Lie 2-group.

Remark 2.4.3. Every Lie 2-group Γ can be obtained from a smooth crossed module. Indeed, one puts $G := \Gamma_0$ and $H := \ker(s)$, equipped with the Lie group structures defined by the multiplication functor m of Γ. The homomorphism $t : H \to G$ is the target map $t : \Gamma_1 \to \Gamma_0$, and the action of G on H is given by the formula ${}^g\gamma := \mathrm{id}_g \cdot \gamma \cdot \mathrm{id}_{g^{-1}}$ for $g \in \Gamma_0$ and $\gamma \in \ker(s)$. These two constructions are inverse to each other (up to canonical Lie group isomorphisms and strict Lie 2-group isomorphisms, respectively).

Example 2.4.4. Consider a connected Lie group H, so that its automorphism group $\mathrm{Aut}(H)$ is again a Lie group [Onishchik and Vinberg 1988]. Then, we have a smooth crossed module $(\mathrm{Aut}(H), H, i, \mathrm{ev})$, where $i : H \to \mathrm{Aut}(H)$ is the assignment of inner automorphisms to group elements, and $\mathrm{ev} : \mathrm{Aut}(H) \times H \to H$ is the evaluation action. The associated Lie 2-group is denoted by $\mathrm{AUT}(H)$ and is called the *automorphism 2-group* of H.

Example 2.4.5. Let

$$1 \longrightarrow H \xrightarrow{t} G \xrightarrow{p} K \longrightarrow 1$$

be an exact sequence of Lie groups, i.e., an exact sequence in which p is a submersion and t is an embedding. The homomorphisms $t : H \to G$ and $p : G \to K$ define action groupoids $G/\!/H$ and $K/\!/G$ as explained in Example 2.1.1. The first one is even a Lie 2-group: the action of G on H is defined by ${}^gh := t^{-1}(gt(h)g^{-1})$. This is well-defined: since

$$p(gt(h)g^{-1}) = p(g)p(t(h))p(g^{-1}) = p(g)p(g)^{-1} = 1,$$

the element $gt(h)g^{-1}$ lies in the image of t, and has a unique preimage. The action is smooth because t is an embedding. The axioms of a crossed module are obviously satisfied.

If a Lie groupoid Γ is a Lie 2-group in virtue of a multiplication functor $m : \Gamma \times \Gamma \to \Gamma$, then the category $\mathcal{B}\mathrm{un}_\Gamma(M)$ of principal Γ-bundles over a smooth manifold M is monoidal:

Definition 2.4.6. Let $P : M_{\mathrm{dis}} \to \Gamma$ and $Q : M_{\mathrm{dis}} \to \Gamma$ be principal Γ-bundles. The *tensor product* $P \otimes Q$ is the anafunctor

$$M_{\mathrm{dis}} \xrightarrow{\mathrm{diag}} M_{\mathrm{dis}} \times M_{\mathrm{dis}} \xrightarrow{P\times Q} \Gamma \times \Gamma \xrightarrow{m} \Gamma.$$

Example 2.4.7. (a) Since trivial principal Γ-bundles $\mathbf{I}_f$ correspond to *smooth* functors $f : M_{\mathrm{dis}} \to \Gamma$ (Example 2.3.3), it is clear that $\mathbf{I}_f \otimes \mathbf{I}_g = \mathbf{I}_{fg}$.

(b) Unwinding Definition 2.4.6 in the general case, the tensor product of two principal Γ-bundles P_1 and P_2 with anchors α_1 and α_2, respectively, and actions ρ_1 and ρ_2, respectively, is given by

$$P_1 \otimes P_2 = \big((P_1 \times_M P_2)\, {}_{m\circ(\alpha_1\times\alpha_2)}\!\times_t \Gamma_1\big) / \sim , \tag{2.4-2}$$

where

$$(p_1, p_2, m(\gamma_1, \gamma_2) \circ \gamma) \sim (\rho_1(p_1, \gamma_1), \rho_2(p_2, \gamma_2), \gamma) \tag{2.4-3}$$

for all $p_1 \in P_1$, $p_2 \in P_2$ and morphisms $\gamma, \gamma_1, \gamma_2 \in \Gamma_1$ satisfying $t(\gamma_i) = \alpha_i(p_i)$ for $i = 1, 2$ and $s(\gamma_1)s(\gamma_2) = t(\gamma)$. The bundle projection is

$$\tilde{\pi}(p_1, p_2, \gamma) := \pi_1(p_1) = \pi_2(p_2),$$

the anchor is $\tilde{\alpha}(p_1, p_2, \gamma) := s(\gamma)$, and the Γ-action is given by

$$\tilde{\rho}((p_1, p_2, \gamma), \gamma') := (p_1, p_2, \gamma \circ \gamma').$$

As a consequence of Lemma 2.3.9 and the fact that the composition of anafunctors is associative up to coherent transformations, we have:

Proposition 2.4.8. *For M a smooth manifold and Γ a Lie 2-group, the tensor product*

$$\otimes : \mathcal{B}\mathrm{un}_\Gamma(M) \times \mathcal{B}\mathrm{un}_\Gamma(M) \to \mathcal{B}\mathrm{un}_\Gamma(M)$$

equips the groupoid of principal Γ-bundles over M with a monoidal structure. Moreover, it turns the stack $\mathcal{B}\mathrm{un}_\Gamma(-)$ into a monoidal stack.

Notice that the tensor unit of the monoidal groupoid $\mathcal{B}\mathrm{un}_\Gamma(M)$ is the trivial principal Γ-bundle $\mathbf{I}_1$ associated to the constant map $1 : M \to \Gamma_0$, or, in terms of anafunctors, the one associated to the constant functor $1 : M \to \Gamma$.

A (weak) *Lie 2-group homomorphism* between Lie 2-groups (Γ, m_Γ) and (Ω, m_Ω) is an anafunctor $\Lambda : \Gamma \to \Omega$ together with a transformation

$$\begin{tikzcd} \Gamma \times \Gamma \arrow[r, "m_\Gamma"] \arrow[d, "\Lambda\times\Lambda"'] & \Gamma \arrow[d, "\Lambda"] \arrow[dl, phantom, "\swarrow\eta"] \\ \Omega \times \Omega \arrow[r, "m_\Omega"'] & \Omega \end{tikzcd} \tag{2.4-4}$$

satisfying the evident coherence condition. Under the equivalence with smooth crossed modules (Remark 2.4.3), Lie 2-group homomorphisms correspond to so-called butterflies [Aldrovandi and Noohi 2009]. A Lie 2-group homomorphism is called *weak equivalence*, if the anafunctor Λ is a weak equivalence. Since extensions and tensor products are both defined via composition of anafunctors, we immediately obtain:

Proposition 2.4.9. *Extension along a Lie 2-group homomorphism $\Lambda : \Gamma \to \Omega$ between Lie 2-groups is a monoidal functor*

$$\Lambda : \mathcal{B}\mathrm{un}_\Gamma(M) \to \mathcal{B}\mathrm{un}_\Omega(M)$$

between monoidal categories. Moreover, these form a monoidal morphism between monoidal stacks.

Since a monoidal functor is an equivalence of monoidal categories if it is an equivalence of the underlying categories, Corollary 2.3.12 implies:

Corollary 2.4.10. *For $\Lambda : \Gamma \to \Omega$ a weak equivalence between Lie 2-groups, the monoidal functor of Proposition 2.4.9 is an equivalence of monoidal categories. Moreover, these form a monoidal equivalence between monoidal stacks.*

If we represent the Lie 2-group Γ by a smooth crossed module $t : H \to G$ as described in Example 2.4.2, we want to determine explicitly what the tensor product looks like under the correspondence of $(G/\!\!/H)$-bundles and principal H-bundles with antiequivariant maps to G; see Lemma 2.2.9.

Lemma 2.4.11. *Let $t : H \to G$ be a crossed module and let P and Q be $G/\!\!/H$-bundles over M. Let (P_H, f) and (Q_H, g) be the principal H-bundles together with their H-antiequivariant maps that belong to P and Q, respectively, under the equivalence of Lemma 2.2.9. Then, the principal H-bundle that corresponds to the tensor product $P \otimes Q$ is given by*

$$(P \otimes Q)_H = (P \times_M Q)/\sim \quad \text{where } (p \star h, q) \sim \big(p, q \star \big({}^{f(p)^{-1}}h\big)\big).$$

The action of H on $(P \otimes Q)_H$ is $[(p,q)] \star h = [(p \star h, q)]$, and the H-antiequivariant map of $(P \otimes Q)_H$ is $[(p,q)] \mapsto f(p) \cdot g(q)$.

Similar to the tensor product of principal Γ-bundles, the dual $P^\vee$ of a principal Γ-bundle P over M is the extension of P along the inversion $i : \Gamma \to \Gamma$ of the 2-group, $P^\vee := i(P)$. The equality $m \circ (\mathrm{id}, i) = 1$ of functors $M \to \Gamma$ induces a *death map* $d : P \otimes P^\vee \to \mathbf{I}_1$. We are going to use this bundle morphism in Section 5.2, but omit a further systematical treatment of duals for the sake of brevity.

3. Version I: groupoid-valued cohomology

We have already mentioned group-valued Čech 1-cocycles in the introduction. They consist of an open cover $\mathcal{U} = \{U_i\}_{i \in I}$ of M and smooth functions $g_{ij} : U_i \cap U_j \to G$ satisfying the cocycle condition $g_{ij} \cdot g_{jk} = g_{ik}$. Segal [1968] realized that this is the same as a smooth functor

$$g : \check{\mathcal{C}}(\mathcal{U}) \to \mathcal{B}G$$

where $\mathcal{B}G$ denotes the one-object groupoid introduced in Example 2.1.1(b) and $\check{\mathcal{C}}(\mathcal{U})$ denotes the *Čech groupoid* corresponding to the cover $\mathcal{U}$. It has objects $\bigsqcup_{i \in I} U_i$ and morphisms $\bigsqcup_{i,j \in I} U_i \cap U_j$, and its structure maps are

$$s(x, i, j) = (x, i), \quad t(x, i, j) = (x, j),$$
$$\mathrm{id}_{(x,i)} = (x, i, i) \quad \text{and} \quad (x, j, k) \circ (x, i, j) = (x, i, k).$$

Analogously, smooth natural transformations between smooth functors $\check{\mathcal{C}}(\mathcal{U}) \to \mathcal{B}G$ give rise to Čech coboundaries. Thus the set $[\check{\mathcal{C}}(\mathcal{U}), \mathcal{B}G]$ of equivalence classes of smooth functors equals the usual first Čech cohomology with respect to the cover $\mathcal{U}$. The classical first Čech-cohomology $\check{\mathrm{H}}^1(M, G)$ of M is hence given by the colimit over all open covers $\mathcal{U}$ of M:

$$\check{\mathrm{H}}^1(M, G) = \varinjlim_{\mathcal{U}} [\check{\mathcal{C}}(\mathcal{U}), \mathcal{B}G].$$

We use this coincidence in order to define the 0-th Čech cohomology with coefficients in a general Lie groupoid Γ:

Definition 3.1. If Γ is a Lie groupoid we set

$$\check{\mathrm{H}}^0(M, \Gamma) := \varinjlim_{\mathcal{U}} [\check{\mathcal{C}}(\mathcal{U}), \Gamma]$$

where the colimit is taken over all covers $\mathcal{U}$ of M and $[\check{\mathcal{C}}(\mathcal{U}), \Gamma]$ denotes the set of equivalence classes of smooth functors $\check{\mathcal{C}}(\mathcal{U}) \to \Gamma$.

Remark 3.2. The choice of the degree is such that $\check{\mathrm{H}}^0(M, \Gamma)$ agrees in the case $\Gamma = G_{\mathrm{dis}}$ (Example 2.1.1(a)) with the classical 0-th Čech-cohomology $\check{\mathrm{H}}^0(M, G)$ of M with values in G.

The geometrical meaning of the set is given in the following well-known theorem, which can be proved, e.g., using Lemma 2.2.8.

Theorem 3.3. *There is a bijection*

$$\check{\mathrm{H}}^0(M, \Gamma) \cong \{\textit{Isomorphism classes of principal } \Gamma\textit{-bundles over } M\}.$$

If Γ is not only a Lie 2-groupoid but a Lie 2-group one can also define a *first* cohomology group $\check{\mathrm{H}}^1(M,\Gamma)$. Indeed, in this case one can consider the Lie 2-groupoid $\mathcal{B}\Gamma$ with one object, morphisms Γ_0 and 2-morphisms Γ_1. Multiplication in Γ gives the composition of morphisms in $\mathcal{B}\Gamma$. Let $[\check{\mathcal{C}}(\mathcal{U}),\mathcal{B}\Gamma]$ denote the set of equivalence classes of smooth, weak 2-functors from the Čech-groupoid $\check{\mathcal{C}}(\mathcal{U})$ to the Lie 2-groupoid $\mathcal{B}\Gamma$. For the definition of weak functors see [Bénabou 1967]. Below we will determine this set explicitly.

Definition 3.4. For a 2-group Γ we set

$$\check{\mathrm{H}}^1(M,\Gamma) := \varinjlim_{\mathcal{U}} [\check{\mathcal{C}}(\mathcal{U}),\mathcal{B}\Gamma].$$

Remark 3.5. This agrees for $\Gamma = G_{\mathrm{dis}}$ with the classical $\check{\mathrm{H}}^1(M,G)$. Furthermore, for an abelian Lie group A the Lie groupoid $\mathcal{B}A$ is even a 2-group and $\check{\mathrm{H}}^1(M,\mathcal{B}A)$ agrees with the classical Čech-cohomology $\check{\mathrm{H}}^2(M,A)$.

Unwinding the above definition, we get Version I of smooth Γ-gerbes:

Definition 3.6. Let Γ be a Lie 2-group, and let $\mathcal{U} = \{U_\alpha\}_{\alpha\in A}$ be an open cover of M.

(1) A Γ*-1-cocycle* with respect to $\mathcal{U}$ is a pair $(f_{\alpha\beta}, g_{\alpha\beta\gamma})$ of smooth maps

$$f_{\alpha\beta} : U_\alpha \cap U_\beta \to \Gamma_0 \quad \text{and} \quad g_{\alpha\beta\gamma} : U_\alpha \cap U_\beta \cap U_\beta \to \Gamma_1$$

satisfying $s(g_{\alpha\beta\gamma}) = f_{\beta\gamma}\cdot f_{\alpha\beta}$ and $t(g_{\alpha\beta\gamma}) = f_{\alpha\gamma}$, and

$$g_{\alpha\beta\delta}\circ(g_{\beta\gamma\delta}\cdot \mathrm{id}_{f_{\alpha\beta}}) = g_{\alpha\gamma\delta}\circ(\mathrm{id}_{f_{\gamma\delta}}\cdot g_{\alpha\beta\gamma}). \tag{3-1}$$

Here, the symbols $\circ$ and $\cdot$ stand for the composition and multiplication of Γ, respectively.

(2) Two Γ-1-cocycles $(f_{\alpha\beta}, g_{\alpha\beta\gamma})$ and $(f'_{\alpha\beta}, g'_{\alpha\beta\gamma})$ are *equivalent*, if there exist smooth maps $h_\alpha : U_\alpha \to \Gamma_0$ and $s_{\alpha\beta} : U_\alpha\cap U_\beta \to \Gamma_1$ with

$$s(s_{\alpha\beta}) = g'_{\alpha\beta}\cdot h_\alpha, \quad t(s_{\alpha\beta}) = h_\beta \cdot g_{\alpha\beta}$$
$$\text{and} \quad (\mathrm{id}_{h_\gamma}\cdot g_{\alpha\beta\gamma})\circ(s_{\beta\gamma}\cdot \mathrm{id}_{f_{\alpha\beta}})\circ(\mathrm{id}_{f_{\beta\gamma}}\cdot s_{\alpha\beta}) = s_{\alpha\gamma}\circ(g'_{\alpha\beta\gamma}\cdot \mathrm{id}_{h_\alpha}).$$

Remark 3.7. For a crossed module $t : H \to G$ and $\Gamma := G/\!\!/H$ the associated Lie 2-group (Example 2.4.2) one can reduce Γ-1-cocycles to pairs

$$\tilde{f}_{\alpha\beta} : U_\alpha\cap U_\beta \to G \quad \text{and} \quad \tilde{g}_{\alpha\beta\gamma} : U_\alpha\cap U_\beta\cap U_\beta \to H,$$

which then satisfies a cocycle condition similar to (3-1). Analogously, coboundaries can be reduced to pairs

$$\tilde{h}_\alpha : U_\alpha \to G \quad \text{and} \quad \tilde{s}_{\alpha\beta} : U_\alpha\cap U_\beta \to H.$$

This yields the common definition of nonabelian cocycles, which can for example be found in [Breen 1990] or [Baez and Stevenson 2009].

Example 3.8. In case of the crossed module $i : H \to \mathrm{Aut}(H)$ with $\Gamma = \mathrm{AUT}(H)$ (see Example 2.4.4) Γ-1-cocycles consist of pairs $\tilde{f}_{\alpha\beta} : U_\alpha \cap U_\beta \to \mathrm{Aut}(H)$ and $\tilde{g}_{\alpha\beta} : U_\alpha \cap U_\beta \cap U_\gamma \to H$. Cocycles of this kind classify so-called Lie groupoid H-extensions [Laurent-Gengoux et al. 2009, Proposition 3.14], which can hence be seen as another equivalent version for $\mathrm{AUT}(H)$-gerbes.

4. Version II: classifying maps

It is well-known that for a Lie group G the *smooth* Čech-cohomology $\check{\mathrm{H}}^1(M, G)$ and the *continuous* Čech-cohomology $\check{\mathrm{H}}^1_c(M, G)$ agree if M is a smooth manifold (in particular paracompact). This can, e.g., be shown by locally approximating continuous cocycles by smooth ones without changing the cohomology class — see [Müller and Wockel 2009] (even for G infinite-dimensional). Below we generalize this fact to nonabelian cohomology for certain Lie 2-groups Γ. Here the continuous Čech-cohomology $\check{\mathrm{H}}^1_c(M, \Gamma)$ is defined in the same way as the smooth one (Definition 3.4) but with all maps continuous instead of smooth. A Lie groupoid Γ is called *smoothly separable*, if the set $\underline{\pi}_0\Gamma$ of isomorphism classes of objects is a smooth manifold for which the projection $\Gamma_0 \to \underline{\pi}_0\Gamma$ is a submersion.

Proposition 4.1. *Let M be a smooth manifold and let Γ be a smoothly separable Lie 2-group. Then, the inclusion*

$$\check{\mathrm{H}}^1(M, \Gamma) \to \check{\mathrm{H}}^1_c(M, \Gamma)$$

of smooth into continuous Čech cohomology is a bijection.

Remark 4.2. It is possible that the assumption of being smoothly separable is not necessary, but a proof not assuming this would certainly be more involved than ours. Anyway, *all* Lie 2-groups we are interested in are smoothly separable.

Proof of Proposition 4.1. We denote by $\underline{\pi}_1\Gamma$ the Lie subgroup of Γ_1 consisting of automorphisms of $1 \in \Gamma_0$. Since it has two commuting group structures — composition and multiplication — it is abelian. The idea of the proof is to reduce the statement via long exact sequences to statements proved in [Müller and Wockel 2009]. The exact sequence we need can be found in [Breen 1990]:

$$\begin{aligned}\check{\mathrm{H}}^0(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^1(M, \mathcal{B}\underline{\pi}_1\Gamma) \to \check{\mathrm{H}}^1(M, \Gamma)&\\ \to \check{\mathrm{H}}^1(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^2(M, \mathcal{B}\underline{\pi}_1\Gamma).&\end{aligned}$$

Note that $\check{\mathrm{H}}^1(M, \Gamma)$ and $\check{\mathrm{H}}^1(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}})$ do not have group structures; hence, exactness is only meant as exactness of pointed sets. But we actually have more structure, namely an action of $\check{\mathrm{H}}^1(M, B\underline{\pi}_1\Gamma)$ on $\check{\mathrm{H}}^1(M,\Gamma)$. This action factors to an action of

$$C := \operatorname{coker}\big(\check{\mathrm{H}}^0(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^1(M, \mathcal{B}\underline{\pi}_1\Gamma)\big).$$

In fact on the nonempty fibers of the morphism $\check{\mathrm{H}}^1(M,\Gamma) \to \check{\mathrm{H}}^1(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}})$ this action is simply transitive. In other words: $\check{\mathrm{H}}^1(M,\Gamma)$ is a C-torsor over

$$K := \ker\big(\check{\mathrm{H}}^1(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^2(M, \mathcal{B}\underline{\pi}_1\Gamma)\big).$$

The same type of sequence also exists in continuous cohomology, namely

$$\check{\mathrm{H}}^0_c(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^1_c(M, \mathcal{B}\underline{\pi}_1\Gamma) \to \check{\mathrm{H}}^1_c(M,\Gamma) \\ \to \check{\mathrm{H}}^1_c(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^2_c(M, \mathcal{B}\underline{\pi}_1\Gamma).$$

With

$$C' := \operatorname{coker}\big(\check{\mathrm{H}}^0_c(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^1_c(M, \mathcal{B}\underline{\pi}_1\Gamma)\big),$$
$$K' := \ker\big(\check{\mathrm{H}}^1_c(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^2_c(M, \mathcal{B}\underline{\pi}_1\Gamma)\big),$$

we exhibit $\check{\mathrm{H}}^1_c(M,\Gamma)$ as a C'-torsor over K'.

The natural inclusions of smooth into continuous cohomology form a chain map between the two sequences. From [Müller and Wockel 2009] we know that they are isomorphisms on the second, fourth and fifth factor. In particular we have an induced isomorphism $K \xrightarrow{\sim} K'$. Lemma 4.3 below additionally shows that the induced morphism $C \to C'$ is an isomorphism. Using these isomorphisms we see that $\check{\mathrm{H}}^1(M,\Gamma)$ and $\check{\mathrm{H}}^1_c(M,\Gamma)$ are both C-torsors over K and that the natural map

$$\check{\mathrm{H}}^1(M,\Gamma) \to \check{\mathrm{H}}^1_c(M,\Gamma)$$

is a morphism of torsors. But each morphism of group torsors is bijective, which concludes the proof. □

Lemma 4.3. *The images of*

$$f : \check{\mathrm{H}}^0(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^1(M, \mathcal{B}\underline{\pi}_1\Gamma) \text{ and } f' : \check{\mathrm{H}}^0_c(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^1_c(M, \mathcal{B}\underline{\pi}_1\Gamma)$$

are isomorphic.

Proof. Recall that $\check{\mathrm{H}}^0(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}})$ is the group of smooth maps $s : M \to \underline{\pi}_0\Gamma$ and $\check{\mathrm{H}}^0_c(M, (\underline{\pi}_0\Gamma)_{\mathrm{dis}})$ is the group of continuous maps $t : M \to \underline{\pi}_0\Gamma$. The groups $\check{\mathrm{H}}^1(M, \mathcal{B}\underline{\pi}_1\Gamma) = \check{\mathrm{H}}^2(M, \underline{\pi}_1\Gamma)$ and $\check{\mathrm{H}}^1_c(M, \mathcal{B}\underline{\pi}_1\Gamma) = \check{\mathrm{H}}^2_c(M, \underline{\pi}_1\Gamma)$ are isomorphic by the result of [Müller and Wockel 2009]. Under the connecting homomorphism

$$\check{\mathrm{H}}^0(\underline{\pi}_0\Gamma, (\underline{\pi}_0\Gamma)_{\mathrm{dis}}) \to \check{\mathrm{H}}^1(\underline{\pi}_0\Gamma, \mathcal{B}\underline{\pi}_1\Gamma)$$

the identity $\mathrm{id}_{\underline{\pi}_0\Gamma}$ is sent to a class ξ_Γ with the property that $f(s) = s^*\xi_\Gamma$ and $f'(t) = t^*\xi_\Gamma$. Hence it suffices to show that for each continuous map $t : M \to \underline{\pi}_0\Gamma$

there is a smooth map $s : M \to \underline{\pi}_0\Gamma$ with $s^*\xi_\Gamma = t^*\xi_\Gamma$. It is well-known that for each continuous map t between smooth manifolds a homotopic smooth map s exists. It remains to show that the pullback $\check{\mathrm{H}}^1(\underline{\pi}_0\Gamma, \mathcal{B}\underline{\pi}_1\Gamma) \to \check{\mathrm{H}}^1(M, \mathcal{B}\underline{\pi}_1\Gamma)$ along smooth maps is homotopy invariant. This can, e.g., be seen by choosing smooth (abelian) $\mathcal{B}\underline{\pi}_1\Gamma$-bundle gerbes as representatives, in which case the homotopy invariance can be deduced from the existence of connections. $\square$

It is a standard result in topology that the continuous G-valued Čech cohomology of paracompact spaces is in bijection with homotopy classes of maps to the classifying space $\mathfrak{B}G$ of the group G. A model for the classifying space $\mathfrak{B}G$ is for example the geometric realization of the nerve of the groupoid $\mathcal{B}G$, or Milnor's join construction [1956].

Now let Γ be a Lie 2-group, and let $|\Gamma|$ denote the geometric realization of the nerve of Γ. Since the nerve is a simplicial topological group, $|\Gamma|$ is a topological group. Version II for smooth Γ-gerbes is this:

Definition 4.4 [Baez and Stevenson 2009]. A classifying map for a smooth Γ-gerbe is a continuous map

$$f : M \to \mathfrak{B}|\Gamma|.$$

We denote by $[M, \mathfrak{B}|\Gamma|]$ the set of homotopy classes of classifying maps.

Proposition 4.5 [Baez and Stevenson 2009, Theorem 1]. *Let Γ be a Lie 2-group. Then there is a bijection*

$$\check{\mathrm{H}}^1_c(M, \Gamma) \cong [M, \mathfrak{B}|\Gamma|]$$

where the topological group $|\Gamma|$ is the geometric realization of the nerve of Γ.

Note that the assumption of [Baez and Stevenson 2009, Theorem 1] that Γ is well-pointed is automatically satisfied because Lie groups are well-pointed. Propositions 4.1 and 4.5 imply the following equivalence theorem between Version I and Version II.

Theorem 4.6. *For M a smooth manifold and Γ a smoothly separable Lie 2-group, there is a bijection*

$$\check{\mathrm{H}}^1(M, \Gamma) \cong [M, \mathfrak{B}|\Gamma|].$$

Remark 4.7. Baez and Stevenson [2009, Section 5.2] argue that the space $\mathfrak{B}|\Gamma|$ is homotopy equivalent to a certain geometric realization of the Lie 2-groupoid $|\mathcal{B}\Gamma|$ from Section 3. Baas, Böstedt and Kro [Baas et al. 2012] have shown that $|\mathcal{B}\Gamma|$ classifies concordance classes of charted Γ-2-bundles. In particular, charted Γ-2-bundles are a further equivalent version of smooth Γ-gerbes.

5. Version III: groupoid bundle gerbes

Several definitions of nonabelian bundle gerbes have appeared in literature so far [Aschieri et al. 2005; Jurčo 2011; Murray et al. 2012]. The approach we give here not only shows a conceptually clear way to define nonabelian bundle gerbes, but also produces systematically a whole bicategory. Moreover, these bicategories form a 2-stack over smooth manifolds (with the Grothendieck topology of surjective submersions).

5.1. ***Definition via the plus construction.*** Recall that the stack $\mathcal{B}\mathrm{un}_\Gamma(-)$ of principal Γ-bundles is monoidal if Γ is a Lie 2-group (Proposition 2.4.8). Associated to the monoidal stack $\mathcal{B}\mathrm{un}_\Gamma(-)$ we have a pre-2-stack

$$\mathcal{T}\mathrm{riv}\,\mathcal{G}\mathrm{rb}_\Gamma(-) := \mathcal{B}(\mathcal{B}\mathrm{un}_\Gamma(-))$$

of *trivial* Γ*-gerbes*. Explicitly, there is one trivial Γ-gerbe $\mathcal{I}$ over every smooth manifold M. The 1-morphisms from $\mathcal{I}$ to $\mathcal{I}$ are principal Γ-bundles over M, and the 2-morphisms between those are morphisms of principal Γ-bundles. Horizontal composition is given by the tensor product of principal Γ-bundles, and vertical composition is the ordinary composition of Γ-bundle morphisms.

Now we apply the *plus construction* of [Nikolaus and Schweigert 2011] in order to stackify this pre-2-stack. The resulting 2-stack is by definition the *2-stack of* Γ*-bundle gerbes*; i.e.,

$$\mathcal{G}\mathrm{rb}_\Gamma(-) := (\mathcal{T}\mathrm{riv}\,\mathcal{G}\mathrm{rb}_\Gamma(-))^+.$$

Unwinding the details of the plus construction, we obtain the following definitions:

Definition 5.1.1. Let M be a smooth manifold. A Γ*-bundle gerbe over* M is a surjective submersion $\pi : Y \to M$, a principal Γ-bundle P over $Y^{[2]}$ and an associative morphism

$$\mu : \pi_{23}^* P \otimes \pi_{12}^* P \to \pi_{13}^* P$$

of Γ-bundles over $Y^{[3]}$.

The morphism μ is called the *bundle gerbe product*. Its associativity is the evident condition for bundle morphisms over $Y^{[4]}$.

In order to proceed with the 1-morphisms, we say that a *common refinement* of two surjective submersions $\pi_1 : Y_1 \to M$ and $\pi_2 : Y_2 \to M$ is a smooth manifold Z together with surjective submersions $Z \to Y_1$ and $Z \to Y_2$ such that the diagram

$$\begin{tikzcd} & Z \arrow[dl] \arrow[dr] & \\ Y_1 \arrow[dr, "\pi_1"'] & & Y_2 \arrow[dl, "\pi_2"] \\ & M & \end{tikzcd}$$

is commutative.

We fix the following convention: suppose P_1 and P_2 are Γ-bundles over surjective submersions U_1 and U_2, respectively, and V is a common refinement of U_1 and U_2. Then, a bundle morphism $\varphi : P_1 \to P_2$ is understood to be a bundle morphism between the pullbacks of P_1 and P_2 to the common refinement V. For example, in the following definition this convention applies to $U_1 = Y_1^{[2]}$, $U_2 = Y_2^{[2]}$ and $V = Z^{[2]}$.

Definition 5.1.2. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be Γ-bundle gerbes over M. A *1-morphism* $\mathcal{A} : \mathcal{G}_1 \to \mathcal{G}_2$ is a common refinement Z of the surjective submersions of $\mathcal{G}_1$ and $\mathcal{G}_2$ together with a principal Γ-bundle Q over Z and a morphism

$$\beta : P_2 \otimes \zeta_1^* Q \to \zeta_2^* Q \otimes P_1$$

of Γ-bundles over $Z^{[2]}$, where $\zeta_1, \zeta_2 : Z^{[2]} \to Z$ are the two projections, such that α is compatible with the bundle gerbe products μ_1 and μ_2.

The compatibility of α with μ_1 and μ_2 means that the diagram

$$\begin{tikzcd}
\pi_{23}^* P_2 \otimes \pi_{12}^* P_2 \otimes \zeta_1^* Q \arrow[rr, "\mu_2 \otimes \mathrm{id}"] \arrow[d, "\mathrm{id} \otimes \zeta_{12}^* \beta"'] & & \pi_{13}^* P_2 \otimes \zeta_1^* Q \arrow[dd, "\zeta_{13}^* \beta"] \\
\pi_{23}^* P_2 \otimes \zeta_2^* Q \otimes \pi_{12}^* P_1 \arrow[d, "\zeta_{23}^* \beta \otimes \mathrm{id}"'] & & \\
\zeta_3^* Q \otimes \pi_{23}^* P_1 \otimes \pi_{12}^* P_1 \arrow[rr, "\mathrm{id} \otimes \mu_1"'] & & \zeta_3^* Q \otimes \pi_{13}^* P_1
\end{tikzcd} \tag{5-1}$$

of morphisms of Γ-bundles over $Z^{[3]}$ is commutative.

If $\mathcal{A}_{12} : \mathcal{G}_1 \to \mathcal{G}_2$ and $\mathcal{A}_{23} : \mathcal{G}_2 \to \mathcal{G}_3$ are 1-morphisms between bundle gerbes over M, the composition $\mathcal{A}_{23} \circ \mathcal{A}_{12} : \mathcal{G}_1 \to \mathcal{G}_3$ is given by the fiber product $Z := Z_{23} \times_{Y_2} Z_{12}$, the principal Γ-bundle $Q := Q_{23} \otimes Q_{12}$ over Z, and the morphism

$$P_3 \otimes \zeta_1^* Q \xrightarrow{\beta_{23} \otimes \mathrm{id}} \zeta_2^* Q_{23} \otimes P_2 \otimes \zeta_1^* Q_{12} \xrightarrow{\mathrm{id} \otimes \beta_{12}} \zeta_2^* Q \otimes P_1.$$

The identity 1-morphism $\mathrm{id}_{\mathcal{G}}$ associated to a Γ-bundle gerbe $\mathcal{G}$ is given by Y regarded as a common refinement of $\pi : Y \to M$ with itself, the trivial Γ-bundle $\mathbf{I}_1$ (the tensor unit of $\mathcal{B}\mathrm{un}_\Gamma(Y)$), and the evident morphism $\mathbf{I}_1 \otimes P \to P \otimes \mathbf{I}_1$.

In order to define 2-morphisms, suppose that $\pi_1 : Y_1 \to M$ and $\pi_2 : Y_2 \to M$ are surjective submersions, and that Z and Z' are common refinements of π_1 and π_2. Let W be a common refinement of Z and Z' with surjective submersions $r : W \to Z$

and $r' : W \to Z'$. We obtain two maps

$$s_1 : W \xrightarrow{r} Z \longrightarrow Y_1 \quad \text{and} \quad t_1 : W \xrightarrow{r'} Z' \longrightarrow Y_1,$$

and, analogously, two maps $s_2, t_2 : W \to Y_2$. These patch together to maps

$$x_W := (s_1, t_1) : W \to Y_1 \times_M Y_1 \quad \text{and} \quad y_W := (s_2, t_2) : W \to Y_2 \times_M Y_2.$$

Definition 5.1.3. Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be Γ-bundle gerbes over M, and let

$$\mathcal{A}, \mathcal{A}' : \mathcal{G}_1 \to \mathcal{G}_2$$

be 1-morphisms. A *2-morphism*

$$\varphi : \mathcal{A} \Rightarrow \mathcal{A}'$$

is a common refinement W of the common refinements Z and Z', together with a morphism

$$\varphi : y_W^* P_2 \otimes r^* Q \to r'^* Q' \otimes x_W^* P_1$$

of Γ-bundles over W that is compatible with the morphisms β and β'.

The compatibility means that a certain diagram over $W^{[2]}$ commutes. Fiberwise over a point $(w, w') \in W \times_M W$ this diagram looks as follows:

(5-2)

$$\begin{tikzcd}
P_2|_{s_2(w'),t_2(w')} \otimes P_2|_{s_2(w),s_2(w')} \otimes Q|_{r(w)} \arrow[r, "\mathrm{id}\otimes\beta"] \arrow[d, "\mu_2\otimes\mathrm{id}"'] & P_2|_{s_2(w'),t_2(w')} \otimes Q|_{r(w')} \otimes P_1|_{s_1(w),s_1(w')} \arrow[d, "\varphi\otimes\mathrm{id}"] \\
P_2|_{s_2(w),t_2(w')} \otimes Q|_{r(w)} \arrow[d, "\mu_2^{-1}\otimes\mathrm{id}"'] & Q'|_{r'(w')} \otimes P_1|_{s_1(w'),t_1(w')} \otimes P_1|_{s_1(w),s_1(w')} \arrow[d, "id\otimes\mu_1"] \\
P_2|_{t_2(w),t_2(w')} \otimes P_2|_{s_2(w),t_2(w)} \otimes Q|_{r(w)} \arrow[d, "\mathrm{id}\otimes\varphi"'] & Q'|_{r'(w')} \otimes P_1|_{s_1(w),t_1(w')} \arrow[d, "id\otimes\mu_1^{-1}"] \\
P_2|_{t_2(w),t_2(w')} \otimes Q'|_{r'(w)} \otimes P_1|_{s_1(w),t_1(w)} \arrow[r, "\beta'\otimes\mathrm{id}"'] & Q'|_{r'(w')} \otimes P_1|_{t_1(w),t_1(w')} \otimes P_1|_{s_1(w),t_1(w)}.
\end{tikzcd}$$

Finally we identify two 2-morphisms $(W_1, r_1, r_1', \varphi_1)$ and $(W_2, r_2, r_2', \varphi_2)$ if the pullbacks of φ_1 and φ_2 to $W \times_{Z \times_{\times} Z'} W'$ agree. Explicitly, this condition means that, for all $w_1 \in W_1$ and $w_2 \in W_2$ with $r_1(w_1) = r_2(w_2)$ and $r_1'(w_1) = r_2'(w_2)$, and for all $p_2 \in y_{W_1}^* P_2 = y_{W_2}^* P_2$ and $q \in r_1^* Q = r_2^* Q$, we have $\varphi_1(p_2, q) = \varphi_2(p_2, q)$.

Remark 5.1.4. • In the above situation of a common refinement W of two common refinements Z, Z' of surjective submersions Y_1, Y_2, the diagram

$$
\begin{tikzcd}
 & Z \arrow[dl] \arrow[dr] & \\
Y_1 & W \arrow[u, "r"'] \arrow[d, "r'"] & Y_2 \\
 & Z' \arrow[ul] \arrow[ur] &
\end{tikzcd}
\tag{5-3}
$$

is *not* necessarily commutative. In fact, diagram (5-3) commutes if and only if the two maps $x_W : W \to Y_1 \times_M Y_1$ and $y_W : W \to Y_2 \times_M Y_2$ factor through the diagonal maps $Y_1 \to Y_1 \times_M Y_1$ and $Y_2 \to Y_2 \times_M Y_2$, respectively.

- In the case that a 2-morphism φ is defined on a common refinement Z for which diagram (5-3) *does* commute, Definition 5.1.3 can be simplified. As remarked before, the two maps x_W and y_W factor through the diagonals, over which the bundles P_1 and P_2 have canonical trivializations (see Corollary 5.2.6). Under these trivializations, φ can be identified with a bundle morphism

$$\varphi : Q \to Q'.$$

Furthermore, the compatibility diagram (5-2) simplifies to the diagram

$$
\begin{tikzcd}
P_2 \otimes \eta_1^* Q \arrow[r, "\beta"] \arrow[d, "\mathrm{id}\otimes\eta_1^*\varphi"'] & \eta_2^* Q \otimes P_1 \arrow[d, "\eta_2^*\varphi\otimes\mathrm{id}"] \\
P_2 \otimes \eta_1^* Q' \arrow[r, "\beta'"'] & \eta_2^* Q' \otimes P_1.
\end{tikzcd}
\tag{5-4}
$$

Next we define the vertical composition

$$\varphi_{23} \bullet \varphi_{12} : \mathcal{A}_1 \Rightarrow \mathcal{A}_3$$

of 2-morphisms $\varphi_{12} : \mathcal{A}_1 \Rightarrow \mathcal{A}_2$ and $\varphi_{23} : \mathcal{A}_2 \Rightarrow \mathcal{A}_3$. The refinement is the fiber product $W := W_{12} \times_{Z_2} W_{23}$ of the covers of φ_{12} and φ_{23}. The bundle gerbe products induce isomorphisms

$$x_W^* P_1 \cong x_{W_{23}}^* P_1 \otimes x_{W_{12}}^* P_1 \quad \text{and} \quad y_W^* P_2 \cong y_{W_{23}}^* P_2 \otimes y_{W_{12}}^* P_2$$

over W. Under these identifications, the morphism $y_W^* P_2 \otimes Q_1 \to Q_3 \otimes x_W^* P_1$ for the 2-morphism $\varphi_{23} \bullet \varphi_{12}$ is defined as

$$y_{W_{23}}^* P_2 \otimes y_{W_{12}}^* P_2 \otimes Q_1 \xrightarrow{\mathrm{id}\otimes\varphi_{12}} y_{W_{23}}^* P_2 \otimes Q_2 \otimes x_{W_{12}}^* P_1 \xrightarrow{\varphi_{23}\otimes\mathrm{id}} Q_3 \otimes x_{W_{23}}^* P_1 \otimes x_{W_{12}}^* P_1.$$

The identity for vertical composition is just the identity refinement and the identity morphism. Finally we come to the horizontal composition

$$\varphi_{23} \circ \varphi_{12} : \mathcal{A}_{23} \circ \mathcal{A}_{12} \Rightarrow \mathcal{A}'_{23} \circ \mathcal{A}'_{12}$$

of 2-morphisms $\varphi_{12} : \mathcal{A}_{12} \Rightarrow \mathcal{A}'_{12}$ and $\varphi_{23} : \mathcal{A}_{23} \Rightarrow \mathcal{A}'_{23}$: its refinement W is given by $W_{12} \times_{(Y_2 \times Y_2)} W_{23}$. We look at the three relevant maps $x_W : W \to Y_1 \times_M Y_1$, $y_W : W \to Y_2 \times_M Y_2$ and $z_W : W \to Y_3 \times_M Y_3$. The morphism φ of the 2-morphism $\varphi_{23} \circ \varphi_{12}$ is defined as the composition

$$z_W^* P_3 \otimes Q_{23} \otimes Q_{12} \xrightarrow{\varphi_{23} \otimes \mathrm{id}} Q'_{23} \otimes y_W^* P_2 \otimes Q_{12} \xrightarrow{\mathrm{id} \otimes \varphi_{12}} Q'_{23} \otimes Q'_{12} \otimes x_W^* P_1.$$

It follows from the properties of the plus construction [Nikolaus and Schweigert 2011] that (a) these definitions fit together into a bicategory $\mathcal{G}\mathrm{rb}_\Gamma(M)$, and that (b) these form a pre-2-stack $\mathcal{G}\mathrm{rb}_\Gamma(-)$ over smooth manifolds. That means there are *pullback 2-functors*

$$f^* : \mathcal{G}\mathrm{rb}_\Gamma(N) \to \mathcal{G}\mathrm{rb}_\Gamma(M)$$

associated to smooth maps $f : M \to N$, and that these are compatible with the composition of smooth maps. Pullbacks of Γ-bundle gerbes, 1-morphisms, and 2-morphisms are obtained by just taking the pullbacks of all involved data. Finally, the plus construction implies (c):

Theorem 5.1.5 [Nikolaus and Schweigert 2011, Theorem 3.3]. *The pre-2-stack $\mathcal{G}\mathrm{rb}_\Gamma(-)$ of Γ-bundle gerbes is a 2-stack.*

Remark 5.1.6. Every 2-stack over smooth manifolds defines a 2-stack over Lie groupoids [Nikolaus and Schweigert 2011, Proposition 2.8]. This way, our approach produces automatically bicategories $\mathcal{G}\mathrm{rb}_\Gamma(\mathcal{X})$ of Γ-bundle gerbes over a Lie groupoid $\mathcal{X}$. In particular, for an action groupoid $\mathcal{X} = M /\!/ G$ we have a bicategory $\mathcal{G}\mathrm{rb}_\Gamma(M /\!/ G)$ of *G-equivariant Γ-bundle gerbes* over M.

In the remainder of this section we give some examples and describe relations between the definitions given here and existing ones.

Example 5.1.7. Let A be an abelian Lie group, for instance $\mathrm{U}(1)$. Then, $\mathcal{B}A$-bundle gerbes are the same as the well-known A-bundle gerbes [Murray 1996]. For more details see Remark 5.1.10 below.

Example 5.1.8. Let (G, H, t, α) be a smooth crossed module, and let $G /\!/ H$ be the associated action groupoid. Then, a $(G /\!/ H)$-bundle gerbe is the same as a *crossed module bundle gerbe* in the sense of Jurčo [2011]. The equivalence relation of "stably isomorphic" of [Jurčo 2011] is given by "1-isomorphic" in terms of the bicategory constructed here. These coincidences come from the equivalence between $(G /\!/ H)$-bundles and so-called G-H-bundles used in [Jurčo 2011; Aschieri et al. 2005] expressed by Lemma 2.3.10. In particular, in case of the automorphism 2-group $\mathrm{AUT}(H)$ of a connected Lie group H, a $\mathrm{AUT}(H)$-bundle gerbe is the same as a H-bibundle gerbe in the sense of [Aschieri et al. 2005].

Example 5.1.9. Let G be a Lie group, so that G_{dis} is a Lie 2-group. Then, there is an equivalence of 2-categories

$$\mathcal{G}\mathrm{rb}_{G_{\mathrm{dis}}}(M) \cong \mathcal{B}\mathrm{un}_{\mathcal{B}G}(M)_{\mathrm{dis}}.$$

Indeed, if $\mathcal{G}$ is a G_{dis}-bundle gerbe over M, its principal G_{dis}-bundle over $Y^{[2]}$ is by Example 2.2.4 just a smooth map $\alpha : Y^{[2]} \to G$, and its bundle gerbe product degenerates to an equality $\pi_{23}^*\alpha \cdot \pi_{12}^*\alpha = \pi_{13}^*\alpha$ for functions on $Y^{[3]}$. In other words, a G_{dis}-bundle gerbe is the same as a so-called G-bundle 0-gerbe. These form a category that is equivalent to the one of ordinary principal G-bundles, as pointed out in Section 1.

Remark 5.1.10. There are two differences between the definitions given here (for $\Gamma = \mathcal{B}A$) and the ones discussed in the list below. Firstly, we have a slightly different ordering of tensor products of bundles. These orderings are not essential in the case of abelian groups because the tensor category of ordinary A-bundles is symmetric. In the nonabelian case, a consistent theory requires the conventions we have chosen here. Secondly, the definitions of 1-morphisms and 2-morphisms have been generalized step by step:

(1) In [Murray 1996], 1-morphisms did not include a *common* refinement, but rather required that the surjective submersion of one bundle gerbe refines the other. This definition is too restrictive in the sense that, e.g., U(1)-bundle gerbes are not classified by $\mathrm{H}^3(M, \mathbb{Z})$, as intended.

(2) In [Murray and Stevenson 2000], 1-morphisms were defined on the canonical refinement $Z := Y_1 \times_M Y_2$ of the surjective submersions of the bundle gerbes. This definition solves the previous problems concerning the classification of bundle gerbes, but makes the composition of 1-morphisms quite involved [Stevenson 2000].

(3) In [Waldorf 2007], 1-morphisms were defined on refinements $\zeta : Z \to Y_1 \times_M Y_2$. This generalization allows the same elegant definition of composition we have given here, and results in the same isomorphism classes of bundle gerbes. Moreover, 2-morphisms are defined with *commutative* diagrams (5-3) — this makes the structure of the bicategory outmost simple (see Remark 5.1.4).

(4) In the present article we have allowed for a yet more general refinement in the definition of 1-morphisms. Its achievement is that bundle gerbes come out as an example of a more general concept — the plus construction — and we get, e.g., Theorem 5.1.5 for free.

Despite these different definitions of 1-morphisms and 2-morphisms, the resulting bicategories of $\mathcal{B}A$-bundle gerbes in (2), (3) and (4) are all equivalent (see [Waldorf

2007, Theorem 1; Nikolaus and Schweigert 2011, Remark 4.5] and Lemma 5.2.8 below).

5.2. ***Properties of groupoid bundle gerbes.*** We recall that a homomorphism $\Lambda : \Gamma \to \Omega$ between Lie 2-groups is an anafunctor together with a transformation (2.4-4) describing its compatibility with the multiplications. We recall further from Proposition 2.4.9 that extension along Λ is a 1-morphism

$$\Lambda : \mathcal{B}\mathrm{un}_{\Gamma}(-) \to \mathcal{B}\mathrm{un}_{\Omega}(-)$$

between monoidal stacks over smooth manifolds. That is, extension along Λ is compatible with pullbacks, tensor products, and morphisms between principal Γ-bundles. Applying it to the principal Γ-bundle P of a Γ-bundle gerbe $\mathcal{G}$, and also to the bundle gerbe product μ, we obtain immediately an Ω-bundle gerbe $\Lambda\mathcal{G}$. The same is evidently true for morphisms and 2-morphisms. Summarizing, we get:

Proposition 5.2.1. *Extension of bundle gerbes along a homomorphism $\Lambda : \Gamma \to \Omega$ between Lie 2-groups defines a 1-morphism*

$$\Lambda : \mathcal{G}\mathrm{rb}_{\Gamma}(-) \to \mathcal{G}\mathrm{rb}_{\Omega}(-)$$

of 2-stacks over smooth manifolds.

We recall that a weak equivalence between Lie 2-groups is a homomorphism $\Lambda : \Gamma \to \Omega$ that is a weak equivalence (see Definition 2.3.11). We have:

Theorem 5.2.2. *Suppose $\Lambda : \Gamma \to \Omega$ is a weak equivalence between Lie 2-groups. Then, the 1-morphism $\Lambda : \mathcal{G}\mathrm{rb}_{\Gamma}(-) \to \mathcal{G}\mathrm{rb}_{\Omega}(-)$ of Proposition 5.2.1 is an equivalence of 2-stacks.*

Proof. The monoidal equivalence $\Lambda : \mathcal{B}\mathrm{un}_{\Gamma}(-) \to \mathcal{B}\mathrm{un}_{\Omega}(-)$ between the monoidal stacks (Corollary 2.4.10) induces an equivalence $\mathcal{T}\mathrm{riv}\,\mathcal{G}\mathrm{rb}_{\Gamma}(M) \to \mathcal{T}\mathrm{riv}\,\mathcal{G}\mathrm{rb}_{\Lambda}(M)$ between pre-2-stacks. Since the plus construction is functorial, this induces in turn the claimed equivalence of 2-stacks. $\square$

Next we generalize a couple of well-known results from abelian to nonabelian bundle gerbes. We define a *refinement* of a surjective submersion $\pi : Y \to M$ to be another surjective submersion $\omega : W \to M$ together with a smooth map $f : W \to Y$ such that $\zeta = \pi \circ f$. Notice that such a refinement induces smooth maps $f_k : W^{[k]} \to Y^{[k]}$ that commute with the various projections $\omega_{i_1 \cdots i_k}$ and $\pi_{i_1 \cdots i_k}$.

Lemma 5.2.3. *Suppose $\mathcal{G}_1 = (Y_1, P_1, \mu_1)$ and $\mathcal{G}_2 = (Y_2, P_2, \mu_2)$ are Γ-bundle gerbes over M, $f : Y_1 \to Y_2$ is a refinement of surjective submersions, and $\varphi : f_2^* P_2 \to P_1$ is an isomorphism of Γ-bundles over $Y_1^{[2]}$ that is compatible with* *the bundle gerbe products μ_1 and μ_2 in the sense that the diagram*

$$\begin{tikzcd} \pi_{23}^* f_2^* P_2 \otimes \pi_{12}^* f_2^* P_2 \arrow[r, "f_3^*\mu"] \arrow[d, "\pi_{23}^*\varphi\otimes\pi_{12}^*\varphi"'] & \pi_{13}^* f_2^* P_2 \arrow[d, "\pi_{13}^*\varphi"] \\ \pi_{23}^* P_1 \otimes \pi_{12}^* P_1 \arrow[r, "\mu"'] & \pi_{13}^* P_1 \end{tikzcd}$$

is commutative. Then, $\mathcal{G}_1$ and $\mathcal{G}_2$ are isomorphic.

The proof works just the same way as in the abelian case: one constructs the 1-isomorphism over the common refinement $Z := Y_1 \times_M Y_2$ in a straightforward way. As a consequence of Lemma 5.2.3 we have:

Proposition 5.2.4. *Let $\mathcal{G} = (Y, P, \mu)$ be a Γ-bundle gerbe over M, and $f : W \to Y$ a refinement of its surjective submersion $\pi : Y \to M$. Then, $(W, f_2^* P, f_3^*\mu)$ is a Γ-bundle gerbe over M, and is isomorphic to $\mathcal{G}$.*

Lemma 5.2.5. *Let $\mathcal{G} = (Y, P, \mu)$ be a Γ-bundle gerbe over M. Then, there exist unique smooth maps $i : P \to P$ and $t : Y \to P$ such that:*

(i) *The diagrams*

$$\begin{tikzcd} P \arrow[r, "i"] \arrow[d, "\chi"'] & P \arrow[d, "\chi"] \\ Y^{[2]} \arrow[r, "\text{flip}"'] & Y^{[2]} \end{tikzcd} \quad \textit{and} \quad \begin{tikzcd} & P \arrow[d, "\chi"] \\ Y \arrow[ur, "t"] \arrow[r, "\text{diag}"'] & Y^{[2]} \end{tikzcd}$$

are commutative.

(ii) *The map t is neutral with respect to the bundle gerbe product μ; i.e.,*

$$\mu(t(y_2), p) = p = \mu(p, t(y_1))$$

for all $p \in P$ with $\chi(p) = (y_1, y_2)$.

(iii) *The map i provides inverses with respect to the bundle gerbe product μ; i.e.,*

$$\mu(i(p), p) = t(y_1) \quad \text{and} \quad \mu(p, i(p)) = t(y_2)$$

for all $p \in P$ with $\chi(p) = (y_1, y_2)$.

Moreover, $\alpha(t(y)) = 1$ and $\alpha(i(p)) = \alpha(p)^{-1}$ for all $p \in P$ and $y \in Y$.

Proof. Concerning uniqueness, suppose (t, i) and (t', i') are pairs of maps satisfying (i), (ii) and (iii). Firstly, we have $t'(y) = \mu(t(y), t'(y)) = t(y)$ and so $t = t'$. Then, $\mu(i(p), p) = t(y_1) = t'(y_1) = \mu(i'(p), p)$ implies $i(p) = i'(p)$, and so $i = i'$. In order to see the existence of t and i, denote by $Q := \mathrm{diag}^* P$ the pullback of P to Y, denote by $Q^\vee$ the dual bundle and by $d : Q \otimes Q^\vee \to \mathbf{I}_1$ the death map. Consider the smooth map

$$Y \xrightarrow{s} \mathbf{I}_1 \xrightarrow{d^{-1}} Q \otimes Q^\vee \xrightarrow{\mu^{-1} \otimes \mathrm{id}_{Q^\vee}} Q \otimes Q \otimes Q^\vee \xrightarrow{\mathrm{id} \otimes d} \mathbf{I}_1 \otimes Q \cong Q \xrightarrow{\mathrm{diag}} P$$

where $s : Y \to \mathbf{I}_1$ is the canonical section (see the proof of Lemma 2.2.6). It is straightforward to see that this satisfies the properties of the map t. Since all maps in the above sequence are (anchor-preserving) bundle morphisms, it is clear that $t \circ \alpha = 1$. □

Corollary 5.2.6. *Let $\mathcal{G} = (Y, P, \mu)$ be a Γ-bundle gerbe over M, and let t and i be the unique maps of Lemma 5.2.5. Then,*

(i) *t is a section of* $\mathrm{diag}^* P$, *and defines a trivialization* $\mathrm{diag}^* P \cong \mathbf{I}_1$;

(ii) *i is a bundle isomorphism $i : P^\vee \to \mathrm{flip}^* P$;*

(iii) *$\mathcal{C}_0 := Y$ and $\mathcal{C}_1 := P$ define a Lie groupoid with source and target maps $\pi_1 \circ \chi$ and $\pi_2 \circ \chi$, respectively, composition μ, identity t and inversion i.*

The following statement is well-known for abelian gerbes; the general version can be proved by a straightforward generalization of the constructions given in the proof of [Waldorf 2007, Proposition 3].

Lemma 5.2.7. *Every 1-morphism $\mathcal{A} : \mathcal{G} \to \mathcal{H}$ between Γ-bundle gerbes over M is invertible.*

The last statement of this section shows a way to bring 1-morphisms and 2-morphisms into a simpler form (see Remark 5.1.10). For bundle gerbes $\mathcal{G}_1$ and $\mathcal{G}_2$ with surjective submersions $\pi_1 : Y_1 \to M$ and $\pi_2 : Y_2 \to M$ we denote by $\mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)$ the Hom-category in the bicategory $\mathcal{G}\mathrm{rb}_\Gamma(M)$, and by $\mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)^{\mathrm{FP}}$ the category whose objects are those 1-morphisms whose common refinement is $Z := Y_1 \times_M Y_2$, and whose 2-morphisms are those 2-morphisms whose refinement is $W := Y_1 \times_M Y_2$ with the maps $r, r' : W \to Z$ the identity maps.

Lemma 5.2.8. *The inclusion $\mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)^{\mathrm{FP}} \to \mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)$ is an equivalence of categories.*

Proof. First we show that it is essentially surjective. We assume $\mathcal{A} : \mathcal{G}_1 \to \mathcal{G}_2$ is a general 1-morphism with a principal Γ-bundle Q over a common refinement Z of the surjective submersions $\pi_1 : Y_1 \to M$ and $\pi_2 : Y_2 \to M$ of the two bundle gerbes. We look at the principal Γ-bundle

$$\tilde{Q} := \kappa_2^* P_2 \otimes \mathrm{pr}_2^* Q \otimes \kappa_1^* P_1$$

over $\tilde{Z} := Y_1 \times_M Z \times_M Y_2$, where

$$\kappa_1 : \tilde{Z} \to Y_1^{[2]} : (y_1, z, y_2) \mapsto (y_1, y_1(z)), \quad \kappa_2 : \tilde{Z} \to Y_2^{[2]} : (y_1, z, y_2) \mapsto (y_2(z), y_2).$$

The projection $\mathrm{pr}_{13} : \tilde{Z} \to Y_1 \times_M Y_2$ is a surjective submersion, and over $\tilde{Z} \times_{Y_1 \times_M Y_2} \tilde{Z}$ we have a bundle morphism $\alpha : \mathrm{pr}_1^* \tilde{Q} \to \mathrm{pr}_2^* \tilde{Q}$ defined over a point $(\tilde{z}, \tilde{z}')$ with $\tilde{z} = (y_1, z, y_2)$ and $\tilde{z}' = (y_1, z', y_2)$ by

$$\begin{tikzcd}
\tilde{Q}_{\tilde{z}} \arrow[r, equal] & P_2|_{y_2(z), y_2} \otimes Q_z \otimes P_1|_{y_1, y_1(z)} \arrow[d, "\mu_2^{-1} \otimes \mathrm{id} \otimes \mathrm{id}"] & \\
& P_2|_{y_2(z'), y_2} \otimes P_2|_{y_2(z), y_2(z')} \otimes Q_z \otimes P_1|_{y_1, y_1(z)} \arrow[d, "\mathrm{id} \otimes \beta \otimes \mathrm{id}"] & \\
& P_2|_{y_2(z'), y_2} \otimes Q_{z'} \otimes P_1|_{y_1(z), y_1(z')} \otimes P_1|_{y_1, y_1(z)} \arrow[d, "\mathrm{id} \otimes \mathrm{id} \otimes \mu_1"] & \\
& P_2|_{y_2(z'), y_2} \otimes Q_{z'} \otimes P_1|_{y_1, y_1(z')} \arrow[r, equal] & \tilde{Q}_{\tilde{z}'}.
\end{tikzcd}$$

The compatibility condition (5-1) implies a cocycle condition for α over the threefold fiber product of $\tilde{Z}$ over $Y_1 \times_M Y_2$, and since principal Γ-bundles form a stack, the pair $(\tilde{Q}, \alpha)$ defines a principal Γ-bundle Q^{FP} over $Z^{\mathrm{FP}} := Y_1 \times_M Y_2$. It is now straightforward to show that the bundle isomorphism β itself descends to a bundle isomorphism β^{FP} over $Z^{\mathrm{FP}} \times_M Z^{\mathrm{FP}}$ in such a way that the triple $(Z^{\mathrm{FP}}, Q^{\mathrm{FP}}, \beta^{\mathrm{FP}})$ forms a 1-morphism $\mathcal{A}^{\mathrm{FP}} : \mathcal{G}_1 \to \mathcal{G}_2$.

In order to show that $\mathcal{A}^{\mathrm{FP}}$ is an essential preimage of $\mathcal{A}$, it remains to construct a 2-morphism $\varphi_{\mathcal{A}}^{\mathrm{FP}} : \mathcal{A} \Rightarrow \mathcal{A}^{\mathrm{FP}}$. In the terminology of Definition 5.1.3, we choose $W = \tilde{Z}$ with $r := \mathrm{pr}_2 : W \to Z$ and $r' := \mathrm{pr}_{13} : W \to Z^{\mathrm{FP}}$. Note that diagram (5-3) does not commute. The maps $x_W : W \to Y_1^{[2]}$ and $y_W : W \to Y_2^{[2]}$ are given by $x_W = s \circ \kappa_1$ and $y_W = \kappa_2$, where $s : Y_1^{[2]} \to Y_1^{[2]}$ switches the factors. Now, the bundle isomorphism of the 2-morphism $\varphi_{\mathcal{A}}^{\mathrm{FP}}$ we want to construct is a bundle isomorphism

$$\varphi : y_W^* P_2 \otimes r^* Q \to \tilde{Q} \otimes x_W^* P_1$$

over W, and is fiberwise over a point $w = (y_1, z, y_2)$ given by

$$\begin{tikzcd}
P_2|_{y_2(z), y_2} \otimes Q_z \arrow[r, "\mathrm{id} \otimes \mathrm{id} \otimes t^{-1}"] & P_2|_{y_2(z), y_2} \otimes Q_z \otimes P_{y_1(z), y_1(z)} \arrow[d, "\mathrm{id} \otimes \mathrm{id} \otimes \mu_1^{-1}"] & \\
& P_2|_{y_2(z), y_2} \otimes Q_z \otimes P_1|_{y_1, y_1(z)} \otimes P_1|_{y_1(z), y_1} \arrow[r, equal] & \tilde{Q}_w \otimes P_1|_{s(y_1, y_1(z))},
\end{tikzcd}$$

where t is the trivialization of $\mathrm{diag}^* P$ of Corollary 5.2.6. The compatibility condition (5-2) is straightforward to check.

Now we show that the inclusion $\mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)^{\mathrm{FP}} \to \mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)$ is full and faithful. Since it is clearly faithful, it only remains to show that it is full. Given a morphism $\mathcal{A} \to \mathcal{A}'$ in $\mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)$, i.e., a common refinement W of $Y_1 \times_M Y_2$ with itself and a bundle morphism φ, we have to find a morphism in $\mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)^{\mathrm{FP}}$ such that the two morphisms are identified under the equivalence relation on bundle gerbe

2-morphisms. We denote the bundles over $Y_1 \times_M Y_2$ corresponding to $\mathcal{A}$ and $\mathcal{A}'$ by Q and Q'. The refinement maps are denoted as before by $r = (s_1, s_2) : W \to Y_1 \times_M Y_2$ and $r' = (t_1, t_2) : W \to Y_1 \times_M Y_2$. Then we obtain an isomorphism $r^*Q \to r^*Q'$ fiberwise over a point $w \in W$ by

$$\tag{5-1}
\begin{tikzcd}
Q|_{s_1(w),s_2(w)} \arrow[r, "d^{-1}\otimes\mathrm{id}"] & P_2^\vee|_{s_2(w),t_2(w)} \otimes P_2|_{s_2(w),t_2(w)} \otimes Q|_{s_1(w),s_2(w)} \arrow[d, "\mathrm{id}\otimes\varphi"] & \\
& P_2^\vee|_{s_2(w),t_2(w)} \otimes Q'|_{t_1(w),t_2(w)} \otimes P_1|_{s_1(w),t_1(w)} \arrow[d, "\mathrm{id}\otimes\beta'^{-1}"] & \\
& P_2^\vee|_{s_2(w),t_2(w)} \otimes P_2|_{s_2(w),t_2(w)} \otimes Q'|_{s_1(w),s_2(w)} \arrow[r, "d\otimes\mathrm{id}"] & Q'|_{s_1(w),s_2(w)}
\end{tikzcd}
$$

where $d : P_2^\vee|_{s_2(w),t_2(w)} \otimes P_2|_{s_2(w),t_2(w)} \to \mathbf{I}_1$ is the death map. One can use the compatibility condition for φ to show that this morphism descends to a morphism $\psi : Q \to Q'$ which is a morphism in $\mathcal{H}\mathrm{om}(\mathcal{G}_1, \mathcal{G}_2)^{\mathrm{FP}}$. The two morphisms (W, ψ) and $(Y_1 \times_M Y_2, \varphi)$ are identified if their pullbacks to

$$W \times_{(Y_1 \times_M Y_2 \times_M Y_1 \times_M Y_2)} (Y_1 \times_M Y_2) = \{w \in W \mid r(w) = r'(w)\} =: W_0$$

are equal. On the one side, the map $W_0 \to W$ is the inclusion and the map $W_0 \to Y_1 \times_M Y_2$ is equal to r. The pullback of ψ along r is by construction the map $r^*Q \to r^*Q'$ from (5-1). On the other side, bundles $x_W^* P_1$ and $y_W^* P_2$ over W_0 have canonical trivializations (Corollary 5.2.6(i)) under which φ becomes also equal to the morphism (5-1). □

5.3. ***Classification by Čech cohomology.*** In this section we prove that Versions I (Čech Γ-1-cocycles) and III (Γ-bundle gerbes) are equivalent. For this purpose, we extract a Čech cocycle from a Γ-bundle gerbe $\mathcal{G}$ over M, and prove that this procedure defines a bijection on the level of equivalence classes (Theorem 5.3.2). First we have to ensure the existence of appropriate open covers.

Lemma 5.3.1. *For every Γ-bundle gerbe $\mathcal{G} = (Y, P, \mu)$ over M there exists an open cover $\mathfrak{U} = \{U_i\}_{i \in I}$ of M with sections $\sigma_i : U_i \to Y$, such that the principal Γ-bundles $(\sigma_i \times \sigma_j)^* P$ over $U_i \cap U_j$ are trivializable.*

Proof. One can choose an open cover such that the 2-fold intersections $U_i \cap U_j$ are contractible. Since every Lie 2-group is a crossed module $G /\!\!/ H$ (Remark 2.4.3), and $G /\!\!/ H$-bundles are ordinary H-bundles (Lemma 2.2.9), these admit sections over contractible smooth manifolds. But a section is enough to trivialize the original Γ-bundle (Lemma 2.2.6). □

Let $\mathcal{G}$ be a Γ-bundle gerbe over M, and let $\mathfrak{U} = \{U_i\}_{i \in I}$ be an open cover with the properties of Lemma 5.3.1. We denote by $M_{\mathfrak{U}}$ the disjoint union of all the open sets U_i, and by $\sigma : M_{\mathfrak{U}} \to Y$ the union of the sections σ_i. Then, σ is a refinement of $\pi : Y \to M$, and we have a Γ-bundle gerbe $\mathcal{G}_{\mathfrak{U},\sigma}$ that is isomorphic to $\mathcal{G}$ (Proposition 5.2.4).

The principal Γ-bundle P_{ij} of $\mathcal{G}_{\mathfrak{U},\sigma}$ over the component $U_i \cap U_j$ is by assumption trivializable. Thus there exists a trivialization $t_{ij} : P_{ij} \to \mathbf{I}_{f_{ij}}$ for smooth functions $f_{ij} : U_i \cap U_j \to \Gamma_0$. We define an isomorphism μ_{ijk} between trivial bundles such that the diagram

$$\begin{tikzcd} P_{jk} \otimes P_{ij} \arrow[r, "\mu"] \arrow[d, "t_{jk}\otimes t_{ij}"'] & P_{ik} \arrow[d, "t_{ik}"] \\ \mathbf{I}_{f_{jk}} \otimes \mathbf{I}_{f_{ij}} \arrow[r, "\mu_{ijk}"'] & \mathbf{I}_{f_{ik}} \end{tikzcd}$$

is commutative. Now we are in the situation of Lemma 5.2.3, which implies that the Γ-bundle gerbe $\mathcal{G}_{\mathfrak{U},\sigma,t} := (M_{\mathfrak{U}}, \mathbf{I}_{f_{ij}}, \mu_{ijk})$ is still isomorphic to $\mathcal{G}$.

Combining Lemma 2.2.8 with Example 2.4.7(a), we see that the isomorphisms μ_{ijk} correspond to smooth maps $g_{ijk} : U_i \cap U_j \cap U_k \to \Gamma_1$ such that $s(g_{ijk}) = f_{jk} \cdot f_{ij}$ and $t(g_{ijk}) = f_{ik}$. The associativity condition for μ_{ijk} implies moreover that

$$g_{\alpha\gamma\delta} \circ (g_{\alpha\beta\gamma} \cdot \mathrm{id}_{f_{\gamma\delta}}) = g_{\alpha\beta\delta} \circ (\mathrm{id}_{f_{\alpha\beta}} \cdot g_{\beta\gamma\delta}).$$

Hence, the collection $\{f_{ij}, g_{ijk}\}$ is a Γ-1-cocycle on M with respect to the open cover $\mathfrak{U}$.

Theorem 5.3.2. *Let M be a smooth manifold and let Γ be a Lie 2-group. The above construction defines a bijection*

$$\{\textit{Isomorphism classes of } \Gamma\textit{-bundle gerbes over } M\} \cong \check{\mathrm{H}}^1(M, \Gamma).$$

Proof. We claim that Γ-bundle gerbes $(M_{\mathfrak{U}}, \mathbf{I}_{f_{ij}}, \mu_{ijk})$ and $(M_{\mathcal{V}}, \mathbf{I}_{h_{ij}}, \nu_{ijk})$ are isomorphic if and only if the corresponding Γ-1-cocycles are equivalent. This proves at the same time that the choices of open covers and sections we have made during the construction do not matter, that the resulting map is well-defined on isomorphism classes, and that this map is injective. Surjectivity follows by assigning to a Γ-1-cocycle (f_{ij}, g_{ijk}) with respect to some cover $\mathfrak{U}$ the Γ-bundle gerbe $(M_{\mathfrak{U}}, \mathbf{I}_{f_{ij}}, \mu_{ijk})$ with μ_{ijk} determined by Lemma 2.2.8.

It remains to prove that claim. We assume $\mathcal{A} = (Z, Q, \alpha)$ is a 1-isomorphism between the Γ-bundle gerbes $(M_{\mathfrak{U}}, \mathbf{I}_{f_{ij}}, \mu_{ijk})$ and $(M_{\mathcal{V}}, \mathbf{I}_{h_{ij}}, \nu_{ijk})$. Similarly to Lemma 5.3.1 one can show that there exists a cover $\mathcal{W}$ of M by open sets W_i that refines both $\mathfrak{U}$ and $\mathcal{V}$, and that allows smooth sections $\omega_i : W_i \to Z$ for which the Γ-bundle $\omega_i^* Q$ is trivializable. In the terminology of the above construction, choosing a trivialization $t : \omega^* Q \to \mathbf{I}_{h_i}$ with smooth maps $h_i : W_i \to \Gamma_0$ over $M_{\mathcal{W}}$

converts the isomorphism α into smooth functions $s_{ij} : W_i \cap W_j \to \Gamma_1$ satisfying $s(s_{ij}) = g'_{ij} \cdot h_i$ and $t(s_{ij}) = h_j \cdot g_{ij}$. The compatibility diagram (5-1) implies the remaining condition that makes (h_i, s_{ij}) an equivalence between the Γ-2-cocycles (f_{ij}, g_{ijk}) and (f'_{ij}, g'_{ijk}). □

6. Version IV: principal 2-bundles

The basic idea of a smooth 2-bundle is that it gives for every point x in the base manifold M a Lie groupoid $\mathcal{P}_x$ varying smoothly with x. Numerous different versions have appeared so far in the literature, e.g., [Bartels 2006; Baez and Schreiber 2007; Wockel 2011; Schommer-Pries 2011]. The main objective of *our* version of principal 2-bundles is to make the definition of the objects (i.e., the 2-bundles) *as simple as possible*, while keeping their isomorphism classes in bijection with nonabelian cohomology. Thus, our principal 2-bundles will be defined using *strict* actions of Lie 2-groups on Lie groupoids, and *not* using anafunctors. The necessary "weakness" will be pushed into the definition of 1-morphisms.

6.1. *Definition of principal 2-bundles.* As an important prerequisite for principal 2-bundles we have to discuss actions of Lie 2-groups on Lie groupoids, and equivariant anafunctors.

Definition 6.1.1. Let $\mathcal{P}$ be a Lie groupoid, and let Γ be a Lie 2-group. A *smooth right action of* Γ *on* $\mathcal{P}$ is a smooth functor $R : \mathcal{P} \times \Gamma \to \mathcal{P}$ such that $R(p, 1) = p$ and $R(\rho, \mathrm{id}_1) = \rho$ for all $p \in \mathcal{P}_0$ and $\rho \in \mathcal{P}_1$, and such that the diagram

$$\begin{tikzcd} \mathcal{P} \times \Gamma \times \Gamma \arrow[r, "\mathrm{id} \times m"] \arrow[d, "R \times \mathrm{id}"'] & \mathcal{P} \times \Gamma \arrow[d, "R"] \\ \mathcal{P} \times \Gamma \arrow[r, "m"'] & \mathcal{P} \end{tikzcd}$$

of smooth functors is commutative (strictly, on the nose).

For example, every Lie 2-group acts on itself via multiplication. Note that, due to strict commutativity, one has $R(R(p, g), g^{-1}) = p$ and $R(R(\rho, \gamma), i(\gamma)) = \rho$ for all $g \in \Gamma_0$, $p \in \mathcal{P}_0$, $\gamma \in \Gamma_1$ and $\rho \in \mathcal{P}_1$.

Remark 6.1.2. This definition could be weakened in two steps. First, one could allow a natural transformation in the above diagram instead of commutativity. Secondly, one could allow R to be an anafunctor instead of an ordinary functor. It turns out that for our purposes the above definition is sufficient.

Definition 6.1.3. Let $\mathcal{X}$ and $\mathcal{Y}$ be Lie groupoids with smooth actions (R_1, ρ_1), (R_2, ρ_2) of a Lie 2-group Γ. An *equivariant structure* on an anafunctor $F : \mathcal{X} \to \mathcal{Y}$

is a transformation

$$\begin{tikzcd}
\mathcal{X} \times \Gamma \arrow[r, "R_1"] \arrow[d, "F\times\mathrm{id}"'] & \mathcal{X} \arrow[d, "F"] \arrow[dl, phantom, "\lambda"] \\
\mathcal{Y} \times \Gamma \arrow[r, "R_2"'] & \mathcal{Y}
\end{tikzcd}$$

satisfying the following condition:

$$\begin{tikzcd}[column sep=small]
\mathcal{X}\times\Gamma\times\Gamma \arrow[rr, "\mathrm{id}\times m"] \arrow[dr, "R_1\times\mathrm{id}"] \arrow[ddd, "F\times\mathrm{id}\times\mathrm{id}"'] & & \mathcal{X}\times\Gamma \arrow[dr, "R_1"] & \\
& \mathcal{X}\times\Gamma \arrow[rr, "R_1"] \arrow[dd, "F\times\mathrm{id}"] \arrow[dl, phantom, "\lambda\times\mathrm{id}"] & & \mathcal{X} \arrow[dd, "F"] \arrow[dll, phantom, "\lambda"] \\
& & & \\
\mathcal{Y}\times\Gamma\times\Gamma \arrow[r, "R_2\times\mathrm{id}"'] & \mathcal{Y}\times\Gamma \arrow[rr, "R_2"'] & & \mathcal{Y}
\end{tikzcd}
=
\begin{tikzcd}[column sep=small]
\mathcal{X}\times\Gamma\times\Gamma \arrow[rr, "\mathrm{id}\times m"] \arrow[dd, "F\times\mathrm{id}\times\mathrm{id}"'] & & \mathcal{X}\times\Gamma \arrow[dr, "R_1"] \arrow[dd, "F\times\mathrm{id}"] & \\
& & & \mathcal{X} \arrow[dd, "F"] \arrow[dl, phantom, "\lambda"] \\
\mathcal{Y}\times\Gamma\times\Gamma \arrow[rr, "\mathrm{id}\times m"] \arrow[dr, "R_2\times\mathrm{id}"'] & & \mathcal{Y}\times\Gamma \arrow[dr, "R_2"] & \\
& \mathcal{Y}\times\Gamma \arrow[rr, "R_2"'] & & \mathcal{Y}.
\end{tikzcd}$$

An anafunctor together with a Γ-equivariant structure is called *Γ-equivariant anafunctor*.

In Appendix A we translate this abstract (but evidently correct) definition of equivariance into more concrete terms involving a Γ_1-action on the total space of the anafunctor.

Definition 6.1.4. If $(F, \lambda) : \mathcal{X} \to \mathcal{Y}$ and $(G, \gamma) : \mathcal{X} \to \mathcal{Y}$ are Γ-equivariant anafunctors, a transformation $\eta : F \Rightarrow G$ is called *Γ-equivariant*, if the following equality of transformation holds:

$$\begin{tikzcd}
\mathcal{X} \times \Gamma \arrow[r, "R_1"] \arrow[d, bend right=50, "G\times\mathrm{id}"'] \arrow[d, bend left=50, "F\times\mathrm{id}"] \arrow[d, phantom, "\Leftarrow \eta\times\mathrm{id}"] & \mathcal{X} \arrow[d, "F"] \arrow[dl, phantom, "\lambda"] \\
\mathcal{Y} \times \Gamma \arrow[r, "R_2"'] & \mathcal{Y}
\end{tikzcd}
=
\begin{tikzcd}
\mathcal{X} \times \Gamma \arrow[r, "R_1"] \arrow[d, "G\times\mathrm{id}"'] & \mathcal{X} \arrow[d, bend right=50, "G"'] \arrow[d, bend left=50, "F"] \arrow[d, phantom, "\Leftarrow \eta"] \arrow[dl, phantom, "\gamma"] \\
\mathcal{Y} \times \Gamma \arrow[r, "R_2"'] & \mathcal{Y}.
\end{tikzcd}$$

It follows from abstract nonsense in the bicategory of Lie groupoids, anafunctors and transformations that we have another bicategory with

- objects: Lie groupoids with smooth right Γ-actions;
- 1-morphisms: Γ-equivariant anafunctors;
- 2-morphisms: Γ-equivariant transformations.

We need three further notions for the definition of a principal 2-bundle. Let M be a smooth manifold, and let $\mathcal{P}$ be a Lie groupoid. We say that a smooth functor $\pi : \mathcal{P} \to M_{\mathrm{dis}}$ is a *surjective submersion functor*, if $\pi : \mathcal{P}_0 \to M$ is a surjective submersion. Let $\pi : \mathcal{P} \to M_{\mathrm{dis}}$ be a surjective submersion functor, and let $\mathcal{Q}$ be a Lie groupoid with some smooth functor $\chi : \mathcal{Q} \to M_{\mathrm{dis}}$. Then, the fiber product $\mathcal{P} \times_M \mathcal{Q}$ is defined to be the full subcategory of $\mathcal{P} \times \mathcal{Q}$ over the submanifold $\mathcal{P}_0 \times_M \mathcal{Q}_0 \subset \mathcal{P}_0 \times \mathcal{Q}_0$.

Definition 6.1.5. Let M be a smooth manifold and let Γ be a Lie 2-group.

(a) A *principal* Γ-2-*bundle over* M is a Lie groupoid $\mathcal{P}$, a surjective submersion functor $\pi : \mathcal{P} \to M_{\mathrm{dis}}$, and a smooth right action R of Γ on $\mathcal{P}$ that preserves π, such that the smooth functor

$$\tau := (\mathrm{pr}_1, R) : \mathcal{P} \times \Gamma \to \mathcal{P} \times_M \mathcal{P}$$

is a weak equivalence.

(b) A 1-*morphism* between principal Γ-2-bundles is a Γ-equivariant anafunctor

$$F : \mathcal{P}_1 \to \mathcal{P}_2$$

that respects the surjective submersion functors to M.

(c) A 2-*morphism* between 1-morphisms is a Γ-equivariant transformation between these.

Remark 6.1.6. (a) The condition in Definition 6.1.5(a) that the action R preserves the surjective submersion functor π means that the diagram of functors

$$\begin{tikzcd} \mathcal{P} \times \Gamma \arrow[r, "R"] \arrow[d, "\mathrm{pr}_1"'] & \mathcal{P} \arrow[d, "\pi"] \\ \mathcal{P} \arrow[r, "\pi"'] & M_{\mathrm{dis}} \end{tikzcd}$$

is commutative.

(b) The condition in Definition 6.1.5(b) that the anafunctor F respects the surjective submersion functors means in the first place that there exists a transformation

$$\begin{tikzcd} \mathcal{P}_1 \arrow[rr, "F"] \arrow[dr, "\pi_1"'] & \arrow[d, phantom, "\Leftarrow"] & \mathcal{P}_2 \arrow[dl, "\pi_2"] \\ & M_{\mathrm{dis}}. & \end{tikzcd}$$

However, since the target of the anafunctors π_1 and $\pi_2 \circ F$ is the *discrete* groupoid M_{dis}, the equivalence of Example 2.3.4 applies, and implies that, if such a transformation exists, it is unique. Indeed, it is easy to see that an anafunctor $F : \mathcal{P} \to \mathcal{Q}$ with anchors $\alpha_l : F \to \mathcal{P}_0$ and $\alpha_r : F \to \mathcal{Q}_0$ respects smooth functors $\pi : \mathcal{P} \to M_{\text{dis}}$ and $\chi : \mathcal{Q} \to M_{\text{dis}}$ if and only if $\pi \circ \alpha_l = \chi \circ \alpha_r$.

Example 6.1.7. The *trivial* Γ*-2-bundle* over M is defined by

$$\mathcal{P} := M_{\text{dis}} \times \Gamma, \quad \pi := \text{pr}_1, \quad R := \text{id}_M \times m.$$

Here, the smooth functor τ even has a smooth inverse *functor*. In the following we denote the trivial Γ-2-bundle by $\mathcal{I}$.

Remark 6.1.8. The principal Γ-2-bundles of Definition 6.1.5 are very similar to those of Bartels [2006] and Wockel [2011], in the sense that their fibers are groupoids with a Γ-action. They only differ in the strictness assumptions for the action, and in the formulation of principality. Opposed to that, the *principal 2-group bundles* introduced in [Ginot and Stiénon 2008] are quite different: their fibers are Lie 2-groupoids equipped with a certain Lie 2-groupoid morphism to $B\Gamma$.

6.2. *Properties of principal 2-bundles.* Principal Γ-2-bundles over M form a bicategory denoted $2\text{-}\mathcal{B}\text{un}_\Gamma(M)$. There is an evident pullback 2-functor

$$f^* : 2\text{-}\mathcal{B}\text{un}_\Gamma(N) \to 2\text{-}\mathcal{B}\text{un}_\Gamma(M)$$

associated to smooth maps $f : M \to N$, and these make $2\text{-}\mathcal{B}\text{un}_\Gamma(-)$ a pre-2-stack over smooth manifolds. We deduce the following important two theorems about this pre-2-stack. The first asserts that it actually is a 2-stack:

Theorem 6.2.1. *Principal* Γ*-2-bundles form a 2-stack* $2\text{-}\mathcal{B}\text{un}_\Gamma(-)$ *over smooth manifolds.*

Proof. This follows from Theorem 5.1.5 (Γ-bundle gerbes form a 2-stack) and Theorem 7.0.1 (the equivalence $\mathcal{G}\text{rb}_\Gamma(-) \cong 2\text{-}\mathcal{B}\text{un}_\Gamma(-)$) we prove in Section 7. □

Remark 6.2.2. Similar to Remark 5.1.6, we obtain automatically bicategories $2\text{-}\mathcal{B}\text{un}_\Gamma(\mathcal{X})$ of principal Γ-2-bundles over Lie groupoids $\mathcal{X}$, including bicategories of *equivariant principal* Γ*-2-bundles*.

The second concerns a homomorphism $\Lambda : \Gamma \to \Omega$ of Lie 2-groups, which induces the extension $\Lambda : \mathcal{G}\text{rb}_\Gamma(-) \to \mathcal{G}\text{rb}_\Omega(-)$ between 2-stacks of bundle gerbes (Proposition 5.2.1). Combined with the equivalence $\mathcal{G}\text{rb}_\Gamma(-) \cong 2\text{-}\mathcal{B}\text{un}_\Gamma(-)$ of Theorem 7.0.1, it defines a 1-morphism

$$\Lambda : 2\text{-}\mathcal{B}\text{un}_\Gamma(-) \to 2\text{-}\mathcal{B}\text{un}_\Omega(-)$$

between 2-stacks of principal 2-bundles. Now we get as a direct consequence of Theorem 5.2.2:

Theorem 6.2.3. *If* $\Lambda : \Gamma \to \Omega$ *is a weak equivalence between Lie* 2*-groups, then the* 1*-morphism* $\Lambda : 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(-) \to 2\text{-}\mathcal{B}\mathrm{un}_\Omega(-)$ *is an equivalence of* 2*-stacks.*

A third consequence of the equivalence of Theorem 7.0.1 in combination with Lemma 5.2.7 is

Corollary 6.2.4. *Every* 1*-morphism* $F : \mathcal{P}_1 \to \mathcal{P}_2$ *between principal* Γ*-*2*-bundles over* M *is invertible.*

The following discussion centers around *local trivializability* that is implicitly contained in Definition 6.1.5. A principal Γ-2-bundle that is isomorphic to the trivial Γ-2-bundle $\mathcal{I}$ introduced in Example 6.1.7 is called *trivializable*. A *section* of a principal Γ-2-bundle $\mathcal{P}$ over M is an anafunctor $S : M_{\mathrm{dis}} \to \mathcal{P}$ such that $\pi \circ S = \mathrm{id}_{M_{\mathrm{dis}}}$ (recall that an anafunctor $\pi \circ S : M \to M$ is the same as a smooth map). One can show that every point $x \in M$ has an open neighborhood U together with a section $s : U_{\mathrm{dis}} \to \mathcal{P}|_U$. Such sections can even be chosen to be smooth functors, rather than anafunctors, namely simply as ordinary sections of the surjective submersion $\pi : (\mathcal{P}|_U)_0 \to U_{\mathrm{dis}}$.

Lemma 6.2.5. *A principal* Γ*-*2*-bundle over* M *is trivializable if and only if it has a smooth section.*

Proof. The trivial Γ-2-bundle $\mathcal{I}$ has the section $S(m) := (m, 1)$, where 1 denotes the unit of Γ_0. If $\mathcal{P}$ is trivializable, and $F : \mathcal{I} \to \mathcal{P}$ is an isomorphism, then, $F \circ S$ is a section of $\mathcal{P}$. Conversely, suppose $\mathcal{P}$ has a section $S : M_{\mathrm{dis}} \to \mathcal{P}$. Then, we get the anafunctor

$$\mathcal{I} = M_{\mathrm{dis}} \times \Gamma \xrightarrow{S \times \mathrm{id}} \mathcal{P} \times \Gamma \xrightarrow{R} \mathcal{P}. \tag{6-1}$$

It has an evident Γ-equivariant structure and respects the projections to M. According to Corollary 6.2.4, this is sufficient to have a 1-isomorphism. □

Corollary 6.2.6. *Every principal* Γ*-*2*-bundle is locally trivializable*; *i.e., every point* $x \in M$ *has an open neighborhood* U *and a* 1*-morphism* $T : \mathcal{I} \to \mathcal{P}|_U$.

Remark 6.2.7. In Wockel's version [2011] of principal 2-bundles, local trivializations are required to be smooth functors and to be invertible as smooth functors, rather than allowing anafunctors. This version turns out to be too restrictive in the sense that the resulting bicategory receives no 2-functor from the bicategory $\mathcal{G}\mathrm{rb}_\Gamma(M)$ of Γ-bundle gerbes that would establish an equivalence.

It is also possible to reformulate our definition of principal 2-bundles in terms of local trivializations. This reformulation gives us criteria which might be easier to check than the actual definition, similar to the case of ordinary principal bundles.

Proposition 6.2.8. *Let $\mathcal{P}$ be a Lie groupoid, $\pi : \mathcal{P} \to M_{\mathrm{dis}}$ be a surjective submersion functor, and R be a smooth right action of Γ on $\mathcal{P}$ that preserves π. Suppose every point $x \in M$ has an open neighborhood U together with a Γ-equivariant anafunctor $T : \mathcal{I} \to \mathcal{P}|_U$ that respects the projections. Then, $\pi : \mathcal{P} \to M_{\mathrm{dis}}$ is a principal Γ-2-bundle over M.*

Proof. We only have to prove that the functor τ is a weak equivalence, and we use Theorem 2.3.13. Since all morphisms of $\mathcal{P}$ have source and target in the same fiber of $\pi : \mathcal{P}_0 \to M_{\mathrm{dis}}$, we may check the two conditions of Theorem 2.3.13 locally, i.e., for $\mathcal{P}|_{U_i}$ where U_i is an open cover of M. Using local trivializations $\mathcal{T}_i : \mathcal{I} \to \mathcal{P}|_{U_i}$, the smooth functor τ translates into the smooth functor $(\mathrm{id}, \mathrm{pr}_1, m) : M_{\mathrm{dis}} \times \Gamma \times \Gamma \to (M_{\mathrm{dis}} \times \Gamma) \times_M (M_{\mathrm{dis}} \times \Gamma)$. This functor is an isomorphism of Lie groupoids, and hence essentially surjective and fully faithful. □

7. Equivalence between bundle gerbes and 2-bundles

In this section we show that Versions III and IV of smooth Γ-gerbes are equivalent in the strongest possible sense:

Theorem 7.0.1. *For M a smooth manifold and Γ a Lie 2-group, there is an equivalence of bicategories*

$$\mathcal{G}\mathrm{rb}_\Gamma(M) \cong 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(M)$$

between the bicategories of Γ-bundle gerbes and principal Γ-2-bundles over M. This equivalence is natural in M; i.e., it is an equivalence between pre-2-stacks.

Since the definitions of the bicategories $\mathcal{G}\mathrm{rb}_\Gamma(M)$ and $2\text{-}\mathcal{B}\mathrm{un}_\Gamma(M)$, *and* the above equivalence are all natural in M, we obtain automatically an induced equivalence for the induced bicategories over Lie groupoids (see Remarks 5.1.6 and 6.2.2).

Corollary 7.0.2. *For $\mathcal{X}$ a Lie groupoid and Γ a Lie 2-group, there is an equivalence*

$$\mathcal{G}\mathrm{rb}_\Gamma(\mathcal{X}) \cong 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(\mathcal{X}).$$

The following outlines the proof of Theorem 7.0.1. In Section 7.1 we construct explicitly a 2-functor

$$\mathscr{E}_M : 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(M) \to \mathcal{G}\mathrm{rb}_\Gamma(M).$$

Then we use a general criterion assuring that $\mathscr{E}_M$ is an equivalence of bicategories. This criterion is stated in Lemma B.1: it requires (A) that $\mathscr{E}_M$ is fully faithful on Hom-categories, and (B) to choose certain preimages of objects and 1-morphisms under $\mathscr{E}_M$. Under these circumstances, Lemma B.1 constructs an inverse 2-functor $\mathscr{R}_M$ together with the required pseudonatural transformations assuring that $\mathscr{E}_M$ and $\mathscr{R}_M$ form an equivalence of bicategories. Condition (A) is proved as Lemma 7.1.7 in Section 7.1. The choices (B) are constructed in Section 7.2.

In order to prove that the 2-functors $\mathscr{E}_M$ extend to the claimed equivalence between pre-2-stacks, we use another criterion stated in Lemma B.3. The only additional assumption of Lemma B.3 is that the given 2-functors $\mathscr{E}_M$ form a 1-morphism of pre-2-stacks; this is proved in Proposition 7.1.8. Then, the inverse 2-functors $\mathscr{R}_M$ obtained before automatically form an inverse 1-morphism between pre-2-stacks.

7.1. ***From principal 2-bundles to bundle gerbes.*** In this section we define the 2-functor $\mathscr{E}_M : 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(M) \to \mathcal{G}\mathrm{rb}_\Gamma(M)$.

Definition of $\mathscr{E}_M$ on objects. Let $\mathcal{P}$ be a principal Γ-2-bundle over M, with projection $\pi : \mathcal{P} \to M$ and right action R of Γ on $\mathcal{P}$. The first ingredient of the Γ-bundle gerbe $\mathscr{E}_M(\mathcal{P})$ is the surjective submersion $\pi : \mathcal{P}_0 \to M$. The second ingredient is a principal Γ-bundle P over $\mathcal{P}_0^{[2]}$. We put

$$P := \mathcal{P}_1 \times \Gamma_0.$$

Bundle projection, anchor and Γ-action are given, respectively, by

$$\chi(\rho, g) := (t(\rho), R(s(\rho), g^{-1})), \quad \alpha(\rho, g) := g \quad \text{and} \quad (\rho, g) \circ \gamma := (R(\rho, \mathrm{id}_{g^{-1}} \cdot \gamma), s(\gamma)). \tag{7.1-1}$$

These definitions are motivated by Remark 7.1.2 below.

Lemma 7.1.1. *This defines a principal Γ-bundle over $\mathcal{P}_0^{[2]}$.*

Proof. First we check that $\chi : P \to \mathcal{P}_0^{[2]}$ is a surjective submersion. Since the functor $\tau = (\mathrm{id}, R)$ is a weak equivalence, we know from Theorem 2.3.13 that

$$f : (\mathcal{P}_0 \times \Gamma_0)\, {}_\tau\!\times_{t\times t} \mathcal{P}_1^{[2]} \to \mathcal{P}_0^{[2]} : (p, g, \rho_1, \rho_2) \mapsto (s(\rho_1), s(\rho_2))$$

is a surjective submersion. Now consider the smooth surjective map

$$g : (\mathcal{P}_0 \times \Gamma_0)\, {}_\tau\!\times_{t\times t} \mathcal{P}_1^{[2]} \to \mathcal{P}_1 \times \Gamma_0 : (p, g, \rho_1, \rho_2) \mapsto (\rho_1^{-1} \circ R(\rho_2, \mathrm{id}_{g^{-1}}), g^{-1}).$$

We have $\chi \circ g = f$; thus, χ is a surjective submersion. Next we check that we have defined an action. Suppose $(\rho, g) \in P$ and $\gamma \in \Gamma_1$ such that $\alpha(\rho, g) = g = t(\gamma)$. Then, $\alpha((\rho, g) \circ \gamma) = s(\gamma)$. Moreover, suppose $\gamma_1, \gamma_2 \in \Gamma_1$ with $t(\gamma_1) = g$ and $t(\gamma_2) = s(\gamma_1)$. Then,

$$\begin{aligned}((\rho, g) \circ \gamma_1) \circ \gamma_2 &= (R(\rho, \mathrm{id}_{g^{-1}} \cdot \gamma_1), s(\gamma_1)) \circ \gamma_2 \\ &= \big(R(\rho, \mathrm{id}_{g^{-1}} \cdot \gamma_1 \cdot \mathrm{id}_{s(\gamma_1)^{-1}} \cdot \gamma_2), s(\gamma_2)\big) = (\rho, g) \circ (\gamma_1 \circ \gamma_2),\end{aligned}$$

where we have used that $\gamma_1 \circ \gamma_2 = \gamma_1 \cdot \mathrm{id}_{s(\gamma_1)^{-1}} \cdot \gamma_2$ in any 2-group. It remains to check that the smooth map

$$\tilde{\tau} : P\, {}_\alpha\!\times_t \Gamma_1 \to P\, {}_\chi\!\times_\chi P : ((\rho, g), \gamma) \mapsto ((\rho, g), (\rho, g) \circ \gamma)$$

is a diffeomorphism. For this purpose, we consider the diagram

$$\begin{tikzcd} & \mathcal{P}_1^{[2]} \arrow[d, "s\times t"] \\ (\mathcal{P}_0 \times \Gamma_0) \times (\mathcal{P}_0 \times \Gamma_0) \arrow[r, "\tau\times\tau"'] & \mathcal{P}_0^{[2]} \times \mathcal{P}_0^{[2]} \end{tikzcd} \tag{7.1-2}$$

and claim that (a) $N_1 := P\ {}_\alpha\!\times_t \Gamma_1$ is a pullback of (7.1-2), (b) $N_2 := P\ {}_\chi\!\times_\chi P$ is a pullback of (7.1-2), and (c) the unique map $N_1 \to N_2$ is $\tilde{\tau}$. Thus, $\tilde{\tau}$ is a diffeomorphism.

In order to prove claim (a) we use again that the functor $\tau = (\mathrm{id}, R)$ is a weak equivalence, so that by Theorem 2.3.13 the triple $(\mathcal{P}_1 \times \Gamma_1, \tau, s \times t)$ is a pullback of (7.1-2). We consider the smooth map

$$\xi : N_1 \to \mathcal{P}_1 \times \Gamma_1 : ((\rho, g), \gamma) \mapsto (R(\rho, \mathrm{id}_{g^{-1}}), \gamma)$$

which is a diffeomorphism because $(\rho, \gamma) \mapsto ((R(\rho, \mathrm{id}_{t(\gamma)}), t(\gamma)), \gamma)$ is a smooth map which is inverse to ξ. Thus, putting $f_1 := \tau \circ \xi$ and $g_1 := (s \times t) \circ \xi$ we see that (N_1, f_1, g_1) is a pullback of (7.1-2). In order to prove claim (b), we put

$$\begin{aligned} f_2((\rho_1, g_1), (\rho_2, g_2)) &:= (R(\rho_1, \mathrm{id}_{g_1^{-1}}), \rho_2), \\ g_2((\rho_1, g_1), (\rho_2, g_2)) &:= \big(R(s(\rho), g_1^{-1}), g_2, R(t(\rho_1), g_1^{-1}), g_1\big). \end{aligned}$$

It is straightforward to check that the cone (N_2, f_2, g_2) makes (7.1-2) commutative. The triple (N_2, f_2, g_2) is also universal: in order to see this suppose N' is any smooth manifold with smooth maps $f' : N' \to \mathcal{P}_1^{[2]}$ and $g' : N' \to (\mathcal{P}_0 \times \Gamma_0) \times (\mathcal{P}_0 \times \Gamma_0)$ so that (7.1-2) is commutative. For $n \in N'$, we write $f'(n) = (\rho_1, \rho_2)$ and $g'(n) = (p_1, g_1, p_2, g_2)$. Then, $\sigma(n) := ((R(\rho_1, \mathrm{id}_{g_2^{-1}}), g_2), (\rho_2, g_1))$ defines a smooth map $\sigma : N' \to P\ {}_\chi\!\times_\chi P$. One checks that $f_2 \circ \sigma = f'$ and $g_2 \circ \sigma = g'$, and that σ is the only smooth map satisfying these equations. This proves that (N_2, f_2, g_2) is a pullback. We are left with claim (c). Here one only has to check that $\tau : N_1 \to N_2$ satisfies $f_2 = f_1 \circ \tau$ and $g_2 = g_1 \circ \tau$. $\square$

Remark 7.1.2. The smooth functor $\tau = (\mathrm{id}, R) : \mathcal{P} \times \Gamma \to \mathcal{P} \times_M \mathcal{P}$ is a weak equivalence, and so has a canonical inverse anafunctor τ^{-1} (Remark 2.3.14). The anafunctor

$$\mathcal{P}_0^{[2]} \xrightarrow{\iota} \mathcal{P} \times_M \mathcal{P} \xrightarrow{c} \mathcal{P} \times_M \mathcal{P} \xrightarrow{\tau^{-1}} \mathcal{P} \times \Gamma \xrightarrow{\mathrm{pr}_2} \Gamma,$$

where c is the functor that switches the factors, corresponds to a principal Γ-bundle over $\mathcal{P}_0^{[2]}$ that is canonically isomorphic to the bundle P defined above.

It remains to provide the bundle gerbe product

$$\mu : \pi_{23}^* P \otimes \pi_{12}^* P \to \pi_{13}^* P,$$

which we define by the formula

$$\mu((\rho_{23}, g_{23}), (\rho_{12}, g_{12})) := (\rho_{12} \circ R(\rho_{23}, \mathrm{id}_{g_{12}}), g_{23}g_{12}). \tag{7.1-3}$$

Lemma 7.1.3. *Formula* (7.1-3) *defines an associative isomorphism*

$$\mu : \pi_{23}^* P \otimes \pi_{12}^* P \to \pi_{13}^* P$$

of principal Γ*-bundles over* $\mathcal{P}_0^{[3]}$.

Proof. First of all, we recall from Example 2.4.7(b) that an element in the tensor product $\pi_{23}^* P \otimes \pi_{12}^* P$ is represented by a triple (p_{23}, p_{12}, γ) where $p_{23}, p_{12} \in P$ with $\pi_1(\chi(p_{23})) = \pi_2(\chi(p_{12}))$, and $\alpha(p_{23}) \cdot \alpha(p_{12}) = t(\gamma)$. In (7.1-3) we refer to triples where $\gamma = \mathrm{id}_{g_{23}g_{12}}$, and this definition extends to triples with general $\gamma \in \Gamma_1$ by employing the equivalence relation

$$(p_1, p_2, \gamma) \sim (p_1 \circ (\gamma \cdot \mathrm{id}_{\alpha(p_2)^{-1}}), p_2, \mathrm{id}_{s(\gamma)}). \tag{7.1.4}$$

The complete formula for μ is then

$$\mu((\rho_{23}, g_{23}), (\rho_{12}, g_{12}), \gamma) = (\rho_{12} \circ R(\rho_{23}, \mathrm{id}_{g_{23}^{-1}} \cdot \gamma), s(\gamma)). \tag{7.1.5}$$

Next we check that (7.1.5) is well-defined under the equivalence relation (7.1.4):

$$\begin{aligned}
&\mu(((\rho_{23}, g_{23}), (\rho_{12}, g_{12}), \gamma)) \\
&\qquad = (\rho_{12} \circ R(\rho_{23}, \mathrm{id}_{g_{23}^{-1}} \cdot \gamma), s(\gamma)) \\
&\qquad = \big(\rho_{12} \circ R(\rho_{23} \circ R(\mathrm{id}_{R(s(\rho_{23}), g_{23}^{-1})}, \gamma \cdot \mathrm{id}_{g_{12}^{-1}}), \mathrm{id}_{g_{12}}), s(\gamma)\big) \\
&\qquad = \mu\big((\rho_{23} \circ R(\mathrm{id}_{R(s(\rho_{23}), g_{23}^{-1})}, \gamma \cdot \mathrm{id}_{g_{12}^{-1}}), s(\gamma)g_{12}^{-1}), (\rho_{12}, g_{12}), \mathrm{id}_{s(\gamma)})\big) \\
&\qquad = \mu\big(((\rho_{23}, g_{23}) \circ (\gamma \cdot \mathrm{id}_{g_{12}^{-1}}), (\rho_{12}, g_{12}), \mathrm{id}_{s(\gamma)})\big).
\end{aligned}$$

Now we have shown that μ is a well-defined map from $\pi_{23}^* P \otimes \pi_{12}^* P$ to $\pi_{13}^* P$, and it remains to prove that it is a bundle morphism. Checking that it preserves fibers and anchors is straightforward. It remains to check that (7.1.5) preserves the Γ-action. We calculate

$$\begin{aligned}
&\mu(((\rho_{23}, g_{23}), (\rho_{12}, g_{12}), \gamma) \circ \tilde{\gamma}) \\
&\quad = \mu((\rho_{23}, g_{23}), (\rho_{12}, g_{12}), \gamma \circ \tilde{\gamma}) = \big(\rho_{23} \circ R(\rho_{12}, \mathrm{id}_{g_{12}} \cdot i(\gamma \circ \tilde{\gamma})), s(\tilde{\gamma})\big) \\
&\quad = \big(\rho_{23} \circ R(R(\rho_{12}, \mathrm{id}_{g_{12}}), i(\gamma) \circ i(\tilde{\gamma})), s(\tilde{\gamma})\big) \\
&\quad = \big(\rho_{23} \circ R(R(\rho_{12}, \mathrm{id}_{g_{12}}), i(\gamma)) \circ R(\mathrm{id}_{R(s(\rho_{12}), g)}, i(\tilde{\gamma})), s(\tilde{\gamma})\big) \\
&\quad = \big(\rho_{23} \circ R(\rho_{12}, \mathrm{id}_{g_{12}} \cdot i(\gamma)) \circ R(\mathrm{id}_{R(s(\rho_{12}), g)}, i(\tilde{\gamma})), s(\tilde{\gamma})\big) \\
&\quad = \big(\rho_{23} \circ R(\rho_{12}, \mathrm{id}_{g_{12}} \cdot i(\gamma)), s(\gamma)\big) \circ \tilde{\gamma} = \mu((\rho_{23}, g_{23}), (\rho_{12}, g_{12}), \gamma) \circ \tilde{\gamma}.
\end{aligned}$$

Summarizing, μ is a morphism of Γ-bundles over $\mathcal{P}_0^{[3]}$. The associativity of μ follows directly from the definitions. □

Definition of $\mathscr{E}_M$ on 1-morphisms. We define a 1-morphism $\mathscr{E}_M(F):\mathscr{E}_M(\mathcal{P}) \to \mathscr{E}_M(\mathcal{P}')$ between Γ-bundle gerbes from a 1-morphism $F:\mathcal{P}\to\mathcal{P}'$ between principal Γ-2-bundles. The refinement of the surjective submersions $\pi:\mathcal{P}\to M$ and $\pi':\mathcal{P}'\to M$ is the fiber product $Z := \mathcal{P}_0 \times_M \mathcal{P}'_0$. Its principal Γ-bundle has the total space

$$Q := F \times \Gamma_0,$$

and its projection, anchor and Γ-action are given, respectively, by

$$\chi(f,g) := \big(\alpha_l(f), R(\alpha_r(f), g^{-1})\big), \quad \alpha(f,g) := g \\ \text{and} \quad (f,g)\circ\gamma := (\rho(f, \mathrm{id}_{g^{-1}}\cdot\gamma), s(\gamma)), \tag{7.1.6}$$

where $\rho : F \times \Gamma_1 \to F$ denotes the Γ_1-action on F that comes from the given Γ-equivariant structure on F (see Appendix A).

Lemma 7.1.4. *This defines a principal Γ-bundle Q over Z.*

Proof. We show first that the projection $\chi : Q \to Z$ is a surjective submersion. Since the functor $\tau' : \mathcal{P}'\times\Gamma \to \mathcal{P}\times_M\mathcal{P}$ is a weak equivalence, we have by Theorem 2.3.13 a pullback

$$\begin{tikzcd} X \arrow[r] \arrow[d, "\xi"'] & (\mathcal{P}'_0\times\Gamma_0)\ {}_R{\times}_t\ (\mathcal{P}'_1\times_M\mathcal{P}'_1) \arrow[d, "s\circ\mathrm{pr}_2"] \\ F\ {}_{\pi'\circ\alpha_l(f)}{\times}_{\pi'}\ \mathcal{P}'_0 \arrow[r] & \mathcal{P}'_0\times_M\mathcal{P}'_0 \end{tikzcd}$$

along the bottom map $(f,p')\mapsto(\alpha_r(f),p')$, which is well-defined because the anafunctor F preserves the projections to M (see Remark 6.1.6(b)). In particular, the map ξ is a surjective submersion. It is easy to see that the smooth map

$$k : X \to F\times\Gamma_0 : ((f,p'),(p'_0,g,\rho,\tilde\rho)) \mapsto (f\circ\rho^{-1}\circ R(\tilde\rho,\mathrm{id}_{g^{-1}}), g^{-1})$$

is surjective. Now we consider the commutative diagram

$$\begin{tikzcd} X \arrow[r, "k"] \arrow[d, "\xi"'] & F\times\Gamma_0 \arrow[d, "\chi"] \\ F\ {}_{\pi'\circ\alpha_l(f)}{\times}_{\pi'}\ \mathcal{P}_0 \arrow[r, "\alpha_l\times\mathrm{id}"'] & \mathcal{P}_0\times_M\mathcal{P}'_0. \end{tikzcd}$$

The surjectivity of k and the fact that ξ and $\alpha_l\times\mathrm{id}$ are surjective submersions shows that χ is one, too.

Next, one checks (as in the proof of Lemma 7.1.1) that the Γ-action on Q defined above is well-defined and preserves the projection. Then it remains to check that the smooth map

$$\xi : Q \,{}_{\alpha}\!\times_t \Gamma_1 \to Q \times_{\mathcal{P}_0 \times_M \mathcal{P}'_0} Q : (f, g, \gamma) \mapsto (f, g, \rho(f, \mathrm{id}_{g^{-1}} \cdot \gamma), s(\gamma))$$

is a diffeomorphism. An inverse map is given as follows. Given an element (f_1, g_1, f_2, g_2) on the right-hand side, we have $\alpha_l(f_1) = \alpha_l(f_2)$, so that there exists a unique element $\rho' \in \mathcal{P}'_1$ such that $f_1 \circ \rho' = f_2$. One calculates that (ρ', g_2) and $(\mathrm{id}_{\alpha_r(f_1)}, g_1)$ are elements of the principal Γ-bundle $\mathcal{P}' \times \Gamma_0$ over $\mathcal{P}_0'^{[2]}$ of Lemma 7.1.1. Thus, there exists a unique element $\gamma \in \Gamma_1$ such that $(\rho', g_2) = (\mathrm{id}_{\alpha_r(f_1)}, g_1) \circ \gamma$. Clearly, $t(\gamma) = g_1$ and $s(\gamma) = g_2$, and we have $\rho' = R(\mathrm{id}_{\alpha_r(f_1)}, \mathrm{id}_{g_1^{-1}} \cdot \gamma)$. We define $\xi^{-1}(f_1, g_1, f_2, g_2) := (f_1, g_1, \gamma)$. The calculation that ξ^{-1} is an inverse for ξ uses property (ii) of Definition A.1 for the action ρ, and is left to the reader. $\square$

The next step in the definition of the 1-morphism $\mathscr{E}(F)$ is to define the bundle morphism

$$\beta : P' \otimes \zeta_1^* Q \to \zeta_2^* Q \otimes P$$

over $Z \times_M Z$. We use the notation of Example 2.4.7(b) for elements of tensor products of principal Γ-bundles; in this notation, the morphism β in the fiber over a point $((p_1, p'_1), (p_2, p'_2)) \in Z \times_M Z$ is given by

$$\beta : ((\rho', g'), (f, g), \gamma) \mapsto ((\tilde{f}, g'gh), (\tilde{\rho}, h^{-1}), \gamma),$$

where $h \in \Gamma_0$ and $\tilde{\rho} \in \mathcal{P}'_1$ are chosen such that $s(\tilde{\rho}) = R(p_2, h^{-1})$ and $t(\tilde{\rho}) = p_1$, and

$$\tilde{f} := \rho(\tilde{\rho}^{-1} \circ f \circ R(\rho', \mathrm{id}_g), \mathrm{id}_h). \tag{7.1.7}$$

Lemma 7.1.5. *This defines an isomorphism between principal Γ-bundles.*

Proof. The existence of choices of $\tilde{\rho}$, h follows because the functor $\tau' : \mathcal{P}' \times \Gamma \to \mathcal{P}' \times_M \mathcal{P}'$ is smoothly essentially surjective (Theorem 2.3.13); in particular, one can choose them locally in a smooth way. We claim that the equivalence relation on $\zeta_2^* Q \otimes P$ identifies different choices; thus, we have a well-defined smooth map. In order to prove this claim, we assume other choices $\tilde{\rho}'$, h'. The pairs $(\tilde{\rho}, h^{-1})$ and $(\tilde{\rho}', h'^{-1})$ are elements in the principal Γ-bundle P' over $\mathcal{P}'_0 \times_M \mathcal{P}'_0$ and sit over the same fiber; thus, there exists a unique $\tilde{\gamma} \in \Gamma_1$ such that $(\tilde{\rho}, h^{-1}) \circ \tilde{\gamma} = (\tilde{\rho}', h'^{-1})$; in particular, $R(\tilde{\rho}, \mathrm{id}_h \cdot \tilde{\gamma}) = \tilde{\rho}'$. Now we have

$$\begin{aligned} ((\tilde{f}, g'gh), (\tilde{\rho}, h^{-1}), \gamma) &= \big((\tilde{f}, g'gh), (\tilde{\rho}, h^{-1}), (\mathrm{id}_{t(\gamma)} \cdot i(\tilde{\gamma}) \cdot \tilde{\gamma}) \circ \gamma\big) \\ &\sim \big((\tilde{f}, g'gh) \circ (\mathrm{id}_{t(\gamma)} \cdot i(\tilde{\gamma})), (\tilde{\rho}, h^{-1}) \circ \tilde{\gamma}, \gamma\big) \end{aligned}$$

so that it suffices to calculate

$$\begin{aligned}(\tilde{f}, g'gh) \circ (\mathrm{id}_{t(\gamma)} \cdot i(\tilde{\gamma})) &= \big(\rho(\tilde{f}, \mathrm{id}_{h^{-1}} \cdot i(\tilde{\gamma})), g'gh'\big) \\ &= \big(\rho(\tilde{\rho}^{-1} \circ f \circ R(\rho', \mathrm{id}_g), i(\tilde{\gamma})), g'gh'\big) \\ &= \big(\rho(R(\tilde{\rho}^{-1}, i(\tilde{\gamma}) \cdot \mathrm{id}_{h'^{-1}}) \circ f \circ R(\rho', \mathrm{id}_g), \mathrm{id}_{h'}), g'gh'\big),\end{aligned}$$

where the last step uses the compatibility condition for ρ from Definition A.1(ii). In any 2-group, we have $i(\tilde{\gamma}) \cdot \mathrm{id}_{s(\tilde{\gamma})} = (\mathrm{id}_{t(\tilde{\gamma})^{-1}} \cdot \tilde{\gamma})^{-1}$, in which case the last line is exactly the formula (7.1.7) for the pair $(\tilde{\rho}', h')$.

Next we check that β is well-defined under the equivalence relation on the tensor product $P' \otimes \zeta_1^* Q$. We have

$$x := ((\rho', g'), (f, g), (\gamma_1 \cdot \gamma_2) \circ \gamma) \sim ((\rho', g') \circ \gamma_1, (f, g) \circ \gamma_2, \gamma) =: x'$$

for $\gamma_1, \gamma_2 \in \Gamma_1$ such that $t(\gamma_1) = g'$, $t(\gamma_2) = g$ and $s(\gamma_1)s(\gamma_2) = t(\gamma)$. Taking advantage of the fact that we can make the same choice of $(\tilde{\rho}, h)$ for both representatives x and x', it is straightforward to show that $\beta(x) = \beta(x')$. Finally, it is obvious from the definition of β that it is anchor-preserving and Γ-equivariant. □

In order to show that the triple (Z, Q, β) defines a 1-morphism between bundle gerbes, it remains to verify that the bundle isomorphism β is compatible with the bundle gerbe products μ_1 and μ_2 in the sense of diagram (5-1). This is straightforward to do and left for the reader.

Definition of $\mathscr{E}_M$ on 2-morphisms, compositors and unitors. Let $F_1, F_2 : \mathcal{P} \to \mathcal{P}'$ be 1-morphisms between principal Γ-bundles over M, and let $\eta : F \Rightarrow G$ be a 2-morphism. Between the Γ-bundles Q_1 and Q_2, which live over the same common refinement $Z = \mathcal{P}_0 \times_M \mathcal{P}_0'$, we find immediately the smooth map

$$\eta : Q_1 \to Q_2 : (f_1, g) \mapsto (\eta(f_1), g)$$

which is easily verified to be a bundle morphism. Its compatibility with the bundle morphisms β_1 and β_2 in the sense of the simplified diagram (5-4) is also easy to check. Thus, we have defined a 2-morphism $\mathscr{E}_M(\eta) : \mathscr{E}_M(F_1) \Rightarrow \mathscr{E}_M(F_2)$.

The compositor for 1-morphisms $F_1 : \mathcal{P} \to \mathcal{P}'$ and $F_2 : \mathcal{P}' \to \mathcal{P}''$ is a bundle gerbe 2-morphism

$$c_{F_1,F_2} : \mathscr{E}_M(F_2 \circ F_1) \to \mathscr{E}_M(F_2) \circ \mathscr{E}_M(F_1).$$

Employing the above constructions, the 1-morphism $\mathscr{E}_M(F_2 \circ F_1)$ is defined on the common refinement $Z_{12} := \mathcal{P}_0 \times_M \mathcal{P}_0''$ and has the Γ-bundle $Q_{12} = (F_1 \times_{\mathcal{P}_0'} F_2)/\mathcal{P}_1' \times \Gamma_0$, whereas the 1-morphism $\mathscr{E}_M(F_2) \circ \mathscr{E}_M(F_1)$ is defined on the common refinement $Z := \mathcal{P}_0 \times_M \mathcal{P}_0' \times_M \mathcal{P}_0''$ and has the Γ-bundle $Q_2 \otimes Q_1$ with $Q_k = F_k \times \Gamma_0$. The compositor c_{F_1,F_2} is defined over the refinement Z with the obvious refinement maps $\mathrm{pr}_{13} : Z \to Z_{12}$ and $\mathrm{id} : Z \to Z$ making diagram (5-3) commutative. It is thus a bundle morphism $c_{F_1,F_2} : \mathrm{pr}_{13}^* Q_{12} \to Q_2 \otimes Q_1$. For elements in a tensor product of Γ-bundles we use the notation of Example 2.4.7(b). Then, we define c_{F_1,F_2} by

$$(7.1.8)\quad ((p,p',p''),(f_1,f_2,g)) \mapsto \big((\rho_2(\tilde{\rho}^{-1}\circ f_2, \mathrm{id}_h), gh), (f_1\circ\tilde{\rho}, h^{-1}), \mathrm{id}_g\big),$$

where $h \in \Gamma_0$ and $\tilde{\rho} : R(p', h^{-1}) \to \alpha_r(f_1) = \alpha_l(f_2)$ are chosen in the same way as in the proof of Lemma 7.1.5. The assignment (7.1.8) does not depend on the choices of h and $\tilde{\rho}$, nor on the choice of the representative (f_1, f_2) in $(F_1 \times_{\mathcal{P}_0'} F_2)/\mathcal{P}_1'$. It is obvious that (7.1.8) is anchor-preserving, and its Γ-equivariance can be seen by choosing $(\tilde{\rho}, h)$ in order to compute $c_{F_1,F_2}((p,p',p''),(f_1,f_2,g))$ and $(\tilde{\rho}', h)$ with $\tilde{\rho}' := R(\tilde{\rho}, \mathrm{id}_{g^{-1}}\cdot\gamma^{-1})$ in order to compute $c_{F_1,F_2}(((p,p',p''),(f_1,f_2,g))\circ\gamma)$. In order to complete the construction of the bundle gerbe 2-morphism c_{F_1,F_2} we have to prove that the bundle morphism c_{F_1,F_2} is compatible with the isomorphisms β_{12} of $\mathscr{E}_M(F_2\circ F_1)$ and $(\mathrm{id}\otimes\beta_1)\circ(\beta_2\otimes\mathrm{id})$ of $\mathscr{E}_M(F_2)\circ\mathscr{E}_M(F_1)$ in the sense of diagram (5-4). We start with an element $((\rho'', g''), (f_{12}, g)) \in \mathscr{E}_M(\mathcal{P}'') \otimes \zeta_1^* Q_{12}$, where $f_{12} = (f_1, f_2)$. We have

$$\beta_{12}((\rho'', g''), (f_{12}, g)) = (\widetilde{f_{12}}, g''gh, \tilde{\rho}, h^{-1})$$

upon choosing $(\tilde{\rho}, h)$ as required in the definition of $\mathscr{E}_M(F_2\circ F_1)$. Writing $\widetilde{f_{12}} = (\tilde{f_1}, \tilde{f_2})$ further we have

$$(7.1.9)\quad \begin{aligned}(\zeta_2^* c_{F_1,F_2}\otimes\mathrm{id})&(\widetilde{f_{12}}, g''gh, \tilde{\rho}, h^{-1})\\ &= \big(\rho_2(\tilde{\rho}_2^{-1}\circ\tilde{f_2}, \mathrm{id}_{h_2}), g''ghh_2, \tilde{f_1}\circ\tilde{\rho}_2, h_2^{-1}, \tilde{\rho}, h^{-1}\big)\end{aligned}$$

upon choosing appropriate $(\tilde{\rho}_2, h_2)$ as required in the definition of c_{F_1,F_2}. This is the result of the clockwise composition of diagram (5-4). Counterclockwise, we first get

$$(\mathrm{id}\otimes\zeta_1^* c_{F_1,F_2})((\rho'', g''), (f_{12}, g)) = (\rho'', g'', f'', gh_1, f', h_1^{-1})$$

for choices $(\tilde{\rho}_1, h_1)$, where $f'' := \rho_2(\tilde{\rho}_1^{-1}\circ f_2, \mathrm{id}_{h_1})$ and $f' := f_1\circ\tilde{\rho}_1$. Next we apply the isomorphism β_2 of $\mathscr{E}_M(F_2)$ and get

$$(\beta_2\otimes\mathrm{id})(\rho'', g'', f'', gh_1, f_1', h_1^{-1}) = (\widetilde{f''}, g''ghh_2, \hat{\rho}, \hat{h}^{-1}, f_1', h_1^{-1})$$

where we have used the choices $(\hat{\rho}, \hat{h})$ defined by $\hat{\rho} := R(\tilde{\rho}_1^{-1}, h_1)\circ R(\tilde{\rho}_2, h^{-1}h_1)$ and $\hat{h} := h_1^{-1}hh_2$. The last step is to apply the isomorphism β_1 of $\mathscr{E}_M(F_2)$ which gives

$$(7.1.10)\quad (\mathrm{id}\otimes\beta_1)(\widetilde{f''}, g''ghh_2, \hat{\rho}, \hat{h}^{-1}, f_1', h_1^{-1}) = (\widetilde{f''}, g''ghh_2, \widetilde{f'}, h_2^{-1}, \tilde{\rho}, h^{-1}),$$

where we have used the choices $(\tilde{\rho}, h)$ from above. Comparing (7.1.9) and (7.1.10), we have obvious coincidence in all but the first and third components. For these remaining factors, coincidence follows from the definitions of the various variables.

Finally, we have to construct unitors. The unitor for a principal Γ-2-bundle $\mathcal{P}$ over M is a bundle gerbe 2-morphism

$$u_{\mathcal{P}} : \mathscr{E}_M(\mathrm{id}_{\mathcal{P}}) \Rightarrow \mathrm{id}_{\mathscr{E}_M(\mathcal{P})}.$$

Abstractly, one can associate to $\mathrm{id}_{\mathscr{E}_M(\mathcal{P})}$ the 1-morphism $\mathrm{id}^{\mathrm{FP}}_{\mathscr{E}_M(\mathcal{P})}$ constructed in the proof of Lemma 5.2.8, and then notice that $\mathrm{id}^{\mathrm{FP}}_{\mathscr{E}_M(\mathcal{P})}$ and $\mathscr{E}_M(\mathrm{id}_{\mathcal{P}})$ are canonically 2-isomorphic. In more concrete terms, the unitor $u_{\mathcal{P}}$ has the refinement $W := \mathcal{P}_0^{[3]}$ with the surjective submersions $r := \mathrm{pr}_{12}$ and $r' := \mathrm{pr}_3$ to the refinements $Z = \mathcal{P}_0^{[2]}$ and $Z' = \mathcal{P}_0$ of the 1-morphisms $\mathscr{E}_M(\mathrm{id}_{\mathcal{P}})$ and $\mathrm{id}_{\mathscr{E}_M(\mathcal{P})}$, respectively. The relevant maps x_W and y_W are pr_{13} and pr_{23}, respectively. The principal Γ-bundle of the 1-morphism $\mathrm{id}_{\mathscr{E}_M(\mathcal{P})}$ is the trivial bundle $Q' = \mathbf{I}_1$. We claim that the principal Γ-bundle Q of $\mathscr{E}_M(\mathrm{id}_{\mathcal{P}})$ is the bundle P of the bundle gerbe $\mathscr{E}_M(\mathcal{P})$. Indeed, the formulae (7.1.6) reduce for the identity anafunctor $\mathrm{id}_{\mathcal{P}}$ to those of (7.1-1). Now, the bundle isomorphism of the unitor $u_{\mathcal{P}}$ is

$$y_W^* P \otimes r^* Q = \mathrm{pr}_{23}^* P \otimes \mathrm{pr}_{12}^* P \xrightarrow{\mu} \mathrm{pr}_{13}^* P \cong r'^* Q' \otimes x_W^* P,$$

where μ is the bundle gerbe product of $\mathscr{E}_M(\mathcal{P})$. The commutativity of diagram (5-2) follows from the associativity of μ.

Proposition 7.1.6. *The assignments $\mathscr{E}_M$ for objects, 1-morphisms and 2-morphisms, together with the compositors and unitors defined above, define a 2-functor*

$$\mathscr{E}_M : 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(M) \to \mathcal{G}\mathrm{rb}_\Gamma(M).$$

Proof. A list of axioms for a 2-functor with the same conventions as we use here can be found in [Schreiber and Waldorf 2008, Appendix A]. The first axiom requires that the 2-functor $\mathscr{E}_M$ respects the vertical composition of 2-morphisms — this follows immediately from the definition.

The second axiom requires that the compositors respect the horizontal composition of 2-morphisms. To see this, let $F_1, F_1' : \mathcal{P} \to \mathcal{P}'$ and $F_2, F_2' : \mathcal{P}' \to \mathcal{P}''$ be 1-morphisms between principal Γ-2-bundles, and let $\eta_1 : F_1 \Rightarrow F_1'$ and $\eta_2 : F_2 \Rightarrow F_2'$ be 2-morphisms. Then, the diagram

$$\begin{tikzcd} \mathscr{E}_M(F_2 \circ F_1) \arrow[r, Rightarrow, "\mathscr{E}_M(\eta_1 \circ \eta_2)"] \arrow[d, Rightarrow, "c_{F_1,F_2}"'] & \mathscr{E}_M(F_2' \circ F_1') \arrow[d, Rightarrow, "c_{F_1',F_2'}"] \\ \mathscr{E}_M(F_2) \circ \mathscr{E}_M(F_1) \arrow[r, Rightarrow, "\mathscr{E}_M(\eta_1) \circ \mathscr{E}_M(\eta_2)"'] & \mathscr{E}_M(F_2') \circ \mathscr{E}_M(F_1') \end{tikzcd}$$

has to commute. Indeed, in order to compute c_{F_1,F_2} and $c_{F_1',F_2'}$ one can make the same choice of $(\tilde{\rho}, h)$, because the transformations η and η_2 preserve the anchors.

Then, commutativity follows from the fact that η_1 and η_2 commute with the groupoid actions and the Γ_1-action according to Definition A.1.

The third axiom describes the compatibility of the compositors with the composition of 1-morphisms in the sense that the diagram

$$\begin{tikzcd} \mathscr{E}_M(F_3 \circ F_2 \circ F_1) \arrow[r, Rightarrow, "c_{F_2\circ F_1, F_3}"] \arrow[d, Rightarrow, "c_{F_3\circ F_2, F_1}"'] & \mathscr{E}_M(F_3) \circ \mathscr{E}_M(F_2 \circ F_1) \arrow[d, Rightarrow, "\mathrm{id}\circ c_{F_2,F_1}"] \\ \mathscr{E}_M(F_3 \circ F_2) \circ \mathscr{E}_M(F_1) \arrow[r, Rightarrow, "c_{F_3,F_2}\circ \mathrm{id}"'] & \mathscr{E}_M(F_3) \circ \mathscr{E}_M(F_2) \circ \mathscr{E}_M(F_1) \end{tikzcd}$$

is commutative. In order to verify this, one starts with an element (f_1, f_2, f_3, g) in $\mathscr{E}_M(F_3 \circ F_2 \circ F_1)$. In order to go clockwise, one chooses pairs $(\tilde{\rho}_{12,3}, h_{12,3})$ and $(\tilde{\rho}_{1,2}, h_{1,2})$ and gets from the definitions

$$\mathrm{CW} = \Big(\big(\rho_3(\tilde{\rho}_{12,3}^{-1} \circ f_3, \mathrm{id}_{h_{12,3}}), gh_{12,3}\big), \\ \big(\rho_2(\tilde{\rho}_{1,2}^{-1} \circ f_2 \circ \tilde{\rho}_{12,3}, \mathrm{id}_{h_{1,2}}), h_{12,3}^{-1}h_{1,2}\big), (f_1 \circ \tilde{\rho}_{1,2}, h_{1,2}^{-1})\Big).$$

Counterclockwise, one can choose firstly again the pair $(\tilde{\rho}_{1,2}, h_{1,2})$ and then the pair $(\tilde{\rho}_{2,3}, h_{2,3})$ with $\tilde{\rho}_{2,3} = R(\tilde{\rho}_{12,3}, \mathrm{id}_{h_{1,2}})$ and $h_{2,3} = h_{1,2}^{-1}h_{12,3}$. Then, one gets

$$\mathrm{CCW} = \Big(\big(\rho_3(\tilde{\rho}_{2,3}^{-1} \circ \rho_3(f_3, \mathrm{id}_{h_{1,2}}), \mathrm{id}_{h_{2,3}}), gh_{1,2}h_{2,3}\big), \\ \big(\rho_2(\tilde{\rho}_{1,2}^{-1} \circ f_2, \mathrm{id}_{h_{1,2}}) \circ \tilde{\rho}_{2,3}, h_{2,3}^{-1}\big), (f_1 \circ \tilde{\rho}_{1,2}, h_{1,2}^{-1})\Big),$$

where one has to use formula (A-2) for the Γ_1-action on the composition of equivariant anafunctors. Using the definitions of $h_{2,3}$ and $\tilde{\rho}_{2,3}$ as well as the axiom of Definition A.1(ii) one can show that CW = CCW.

The fourth and last axiom requires that compositors and unitors are compatible with each other in the sense that for each 1-morphism $F : \mathcal{P} \to \mathcal{P}'$ the 2-morphisms

$$\mathscr{E}_M(F) \cong \mathscr{E}_M(F \circ \mathrm{id}_{\mathcal{P}}) \xRightarrow{c_{\mathrm{id}_{\mathcal{P}},F}} \mathscr{E}_M(F) \circ \mathscr{E}_M(\mathrm{id}_{\mathcal{P}}) \xRightarrow{\mathrm{id}\circ u_{\mathcal{P}}} \mathscr{E}_M(F) \circ \mathrm{id}_{\mathscr{E}_M(\mathcal{P})} \cong \mathscr{E}_M(F),$$

$$\mathscr{E}_M(F) \cong \mathscr{E}_M(\mathrm{id}_{\mathcal{P}'} \circ F) \xRightarrow{c_{F,\mathrm{id}_{\mathcal{P}'}}} \mathscr{E}_M(\mathrm{id}_{\mathcal{P}'}) \circ \mathscr{E}_M(F) \xRightarrow{u_{\mathcal{P}'}\circ \mathrm{id}} \mathrm{id}_{\mathscr{E}_M(\mathcal{P}')} \circ \mathscr{E}_M(F) \cong \mathscr{E}_M(F)$$

are the identity 2-morphisms. We prove this for the first one and leave the second as an exercise. Using the definitions, we see that the 2-morphism has the refinement $W := \mathcal{P}_0 \times_M \mathcal{P}_0 \times_M \mathcal{P}_0'$ with $r = \mathrm{pr}_{13}$ and $r' = \mathrm{pr}_{23}$. The maps $x_W : W \to \mathcal{P}_0 \times_M \mathcal{P}_0$ and $y_W : W \to \mathcal{P}_0' \times_M \mathcal{P}_0'$ are pr_{12} and $\Delta \circ \mathrm{pr}_3$, respectively, where Δ is the diagonal map. Its bundle morphism is a morphism

$$\varphi : \mathrm{pr}_{13}^* Q \to \mathrm{pr}_{23}^* Q \otimes \mathrm{pr}_{12}^* P,$$

where $Q = F \times \Gamma_0$ is the principal Γ-bundle of $\mathscr{E}_M(F)$, and $P = \mathcal{P}_1 \times \Gamma_0$ is the principal Γ-bundle of $\mathscr{E}_M(\mathcal{P})$. Over a point (p_1, p_2, p') and $(f, g) \in \mathrm{pr}_{13}^* Q$, i.e.,

$\alpha_l(f) = p_1$ and $R(\alpha_r(f), g^{-1}) = p'$, the bundle morphism φ is given by

$$(f, g) \mapsto \big(\rho(\tilde{\rho}^{-1} \circ f, \mathrm{id}_h), gh, \tilde{\rho}, h^{-1}\big),$$

where $h \in \Gamma_0$, and $\tilde{\rho} \in \mathcal{P}_1$ with $s(\tilde{\rho}) = R(p_2, h^{-1})$ and $t(\tilde{\rho}) = \alpha_l(f)$. We have to compare (W, φ) with the identity 2-morphism of $\mathscr{E}_M(F)$, which has the refinement Z with $r = r' = \mathrm{id}$ and the identity bundle morphism. According to the equivalence relation on bundle gerbe 2-morphisms we have to evaluate φ over a point $w \in W$ with $r(w) = r'(w)$; i.e., w is of the form $w = (p, p, p')$. Here we can choose $h = 1$ and $\tilde{\rho} = \mathrm{id}_p$, in which case we have

$$\varphi(f, g) = ((f, g), (\mathrm{id}_p, 1)).$$

This is indeed the identity on Q. □

Properties of the 2*-functor* $\mathscr{E}_M$. For the proof of Theorem 7.0.1 we provide the following two statements.

Lemma 7.1.7. *The* 2*-functor* $\mathscr{E}_M$ *is fully faithful on Hom-categories.*

Proof. Let $\mathcal{P}$, $\mathcal{P}'$ be principal Γ-2-bundles over M, and let $F_1, F_2 : \mathcal{P} \to \mathcal{P}'$ be 1-morphisms. By Lemma 5.2.8 every 2-morphism $\eta : \mathscr{E}_M(F_1) \Rightarrow \mathscr{E}_M(F_2)$ can be represented by one whose refinement is $\mathcal{P}_0 \times_M \mathcal{P}'_0$, so that its bundle isomorphism is $\eta : Q_1 \to Q_2$, where $Q_k := F_k \times \Gamma$ for $k = 1, 2$. We can read off a map $\eta : F_1 \to F_2$, and it is easy to see that this is a 2-morphism $\eta : F_1 \Rightarrow F_2$. This procedure is clearly inverse to the 2-functor $\mathscr{E}_M$ on 2-morphisms. □

Proposition 7.1.8. *The* 2*-functors* $\mathscr{E}_M$ *form a* 1*-morphism between pre-*2*-stacks.*

Proof. For a smooth map $f : M \to N$, we have to look at the diagram

$$\begin{tikzcd}
2\text{-}\mathcal{B}\mathrm{un}_\Gamma(N) \arrow[r, "f^*"] \arrow[d, "\mathscr{E}_N"'] & 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(M) \arrow[d, "\mathscr{E}_M"] \\
\mathcal{G}\mathrm{rb}_\Gamma(N) \arrow[r, "f^*"'] & \mathcal{G}\mathrm{rb}_\Gamma(M)
\end{tikzcd}$$

of 2-functors. For $\mathcal{P}$ a principal Γ-2-bundle over N, the Γ-bundle gerbe $\mathscr{E}_M(f^*\mathcal{P})$ has the surjective submersion $\mathrm{pr}_1 : Y := M \times_N \mathcal{P}_0 \to M$, the principal Γ-bundle $P := M \times_N \mathcal{P}_1 \times \Gamma_0$ over $Y^{[2]}$, and a bundle gerbe product μ defined as in (7.1-3) that ignores the M-factor. On the other hand, the Γ-bundle gerbe $f^*\mathscr{E}_N(\mathcal{P})$ has the same surjective submersion, and — up to canonical identifications between fiber products — the same Γ-bundle and the same bundle gerbe product. These canonical identifications make up a pseudonatural transformation that renders the above diagram commutative. □

7.2. ***From bundle gerbes to principal 2-bundles.*** We now provide the data we will feed into Lemma B.1 in order to produce a 2-functor $\mathcal{R}_M : \mathcal{G}\mathrm{rb}_\Gamma(M) \to 2\text{-}\mathcal{B}\mathrm{un}_\Gamma(M)$ that is inverse to the 2-functor $\mathcal{E}_M$ constructed in the previous section. These data are:

(1) A principal Γ-2-bundle $\mathcal{R}_\mathcal{G}$ for each Γ-bundle gerbe $\mathcal{G}$ over M.

(2) A 1-isomorphism $\mathcal{A}_\mathcal{G} : \mathcal{G} \to \mathcal{E}_M(\mathcal{R}_\mathcal{G})$ for each Γ-bundle gerbe $\mathcal{G}$ over M.

(3) A 1-isomorphism $\mathcal{R}_\mathcal{A} : \mathcal{P} \to \mathcal{P}'$ and a 2-isomorphism $\eta_\mathcal{A} : \mathcal{A} \Rightarrow \mathcal{E}_M(\mathcal{R}_\mathcal{A})$ for all principal Γ-2-bundles $\mathcal{P}$, $\mathcal{P}'$ over M and all bundle gerbe 1-morphisms $\mathcal{A} : \mathcal{E}_M(\mathcal{P}) \to \mathcal{E}_M(\mathcal{P}')$.

Construction of the principal Γ*-2-bundle* $\mathcal{R}_\mathcal{G}$. We assume that $\mathcal{G}$ consists of a surjective submersion $\pi : Y \to M$, a principal Γ-bundle P over $Y^{[2]}$ and a bundle gerbe product μ. Let $\alpha : P \to \Gamma_0$ be the anchor of P, and let $\chi : P \to Y^{[2]}$ be the bundle projection.

The Lie groupoid $\mathcal{P}$ of the principal 2-bundle $\mathcal{R}_\mathcal{G}$ is defined by

$$\mathcal{P}_0 := Y \times \Gamma_0 \quad \text{and} \quad \mathcal{P}_1 := P \times \Gamma_0;$$

source map, target maps, and composition are given by, respectively,

$$\begin{gathered} s(p, g) := (\pi_2(\chi(p)), g), \quad t(p, g) := \big(\pi_1(\chi(p)), \alpha(p)^{-1} \cdot g\big) \\ \text{and} \quad (p_2, g_2) \circ (p_1, g_1) := (\mu(p_1, p_2), g_1). \end{gathered} \tag{7.2.1}$$

The identity morphism of an object $(y, g) \in \mathcal{P}_0$ is $(t_y, g) \in \mathcal{P}_1$, where t_y denotes the unit element in P over the point (y, y); see Lemma 5.2.5. The inverse of a morphism $(p, g) \in \mathcal{P}_1$ is $(i(p), \alpha(p)^{-1} g)$, where $i : P \to P$ is the map from Lemma 5.2.5. The bundle projection is $\pi(y, g) := \pi(y)$. The action is given on objects and morphisms by

$$\begin{gathered} R_0((y, g), g') := (y, gg'), \\ R_1((p, g), \gamma) := \big(p \circ (\mathrm{id}_g \cdot \gamma \cdot \mathrm{id}_{t(\gamma)^{-1} g^{-1} \alpha(p)}), g \cdot s(\gamma)\big). \end{gathered} \tag{7.2.2}$$

Lemma 7.2.1. *This defines a functor* $R : \mathcal{P} \times \Gamma \to \mathcal{P}$, *and* R *is an action of* Γ *on* $\mathcal{P}$.

Proof. We assume that $t : H \to G$ is a smooth crossed module, and that Γ is the Lie 2-group associated to it; see Example 2.4.2 and Remark 2.4.3. Then we use the correspondence between principal Γ-bundles and principal H-bundles with H-antiequivariant maps to G of Lemma 2.2.9. Writing $\gamma = (h, g')$, we have

$$R_1((p, g), \gamma) = (p \star {}^g h, gg').$$

With this simple formula at hand it is straightforward to show that R respects source and target maps and satisfies the axiom of an action. For the composition, we assume composable $(p_2, g_2), (p_1, g_1) \in \mathcal{P}_1$; i.e., $g_2 = \alpha(p_1)^{-1} g_1$, and composable $(h_2, g_2'), (h_1, g_1') \in \Gamma_1$; i.e., $g_2' = t(h_1) g_1'$. Then we have

$$\begin{aligned}
&R\big((p_2, g_2) \circ (p_1, g_1), (h_2, g_2') \circ (h_1, g_1')\big) \\
&\qquad = R\big((\mu(p_1, p_2), g_1), (h_2 h_1, g_1')\big) = \big(\mu(p_1, p_2) \star {}^{g_1}(h_2 h_1), g_1 g_1'\big) \\
&\qquad = \big(\mu(p_1 \star {}^{g_1}h_2, p_2) \star {}^{g_1}h_1, g_1 g_1'\big) = \big(\mu(p_1, p_2 \star {}^{g_2}h_2) \star {}^{g_1}h_1, g_1 g_1'\big) \\
&\qquad = \big(\mu(p_1 \star {}^{g_1}h_1, p_2 \star {}^{g_2}h_2), g_1 g_1'\big) = (p_2 \star {}^{g_2}h_2, g_2 g_2') \circ (p_1 \star {}^{g_1}h_1, g_1 g_1') \\
&\qquad = R((p_2, g_2), (h_2, g_2')) \circ R((p_1, g_1), (h_1, g_1')),
\end{aligned}$$

finishing the proof. □

It is obvious that the action R preserves the projection π. Thus, in order to complete the construction of the principal 2-bundle $\mathcal{R}_{\mathcal{G}}$ it remains to show that the functor $\tau = (\mathrm{pr}_1, R)$ is a weak equivalence. This is the content of the following two lemmata in connection with Theorem 2.3.13.

Lemma 7.2.2. *The functor τ is smoothly essentially surjective.*

Proof. The condition we have to check is whether or not the map

$$(Y \times \Gamma_0 \times \Gamma_0)\,{}_\tau\!\times_t ((P \times \Gamma_0) \times_M (P \times \Gamma_0)) \xrightarrow{(s \times s) \circ \mathrm{pr}_2} (Y \times \Gamma_0) \times_M (Y \times \Gamma_0)$$

is a surjective submersion. The left-hand side is diffeomorphic to $(P \times \Gamma_0)\,{}_{\pi_1}\!\times_{\pi_1} (P \times \Gamma_0)$ via pr_2, so that this is equivalent to checking that

$$s \times s : (P \times \Gamma_0)\,{}_{\pi_1 \circ \chi}\!\times_{\pi_1 \circ \chi} (P \times \Gamma_0) \to (Y \times \Gamma_0) \times_M (Y \times \Gamma_0)$$

is a surjective submersion. Since the Γ_0-factors are just spectators, this is in turn equivalent to checking that

$$(\pi_2 \times \pi_2) \circ (\chi \times \chi) : P\,{}_{\pi_1 \circ \chi}\!\times_{\pi_1 \circ \chi} P \to Y^{[2]}$$

is a surjective submersion. It fits into the pullback diagram

$$\begin{tikzcd}
P\,{}_{\pi_1 \circ \chi}\!\times_{\pi_1 \circ \chi} P \arrow[r, hook] \arrow[d, "\chi \times \chi"'] & P \times P \arrow[d, "\chi \times \chi"] \\
Y^{[2]}\,{}_{\pi_1}\!\times_{\pi_1} Y^{[2]} \arrow[r, hook] \arrow[d, "\pi_2 \times \pi_2"'] & Y^{[2]} \times Y^{[2]} \arrow[d, "\pi_2 \times \pi_2"] \\
Y^{[2]} \arrow[r, hook] & Y \times Y
\end{tikzcd}$$

which has a surjective submersion on the right-hand side; hence, also the map on the left-hand side must be a surjective submersion. □

Lemma 7.2.3. *The functor τ is smoothly fully faithful.*

Proof. We assume a smooth manifold N with two smooth maps

$$f : N \to (\mathcal{P}_0 \times \Gamma_0) \times (\mathcal{P}_0 \times \Gamma_0) \quad \text{and} \quad g : N \to \mathcal{P}_1 \times_M \mathcal{P}_1$$

such that the diagram

$$\begin{tikzcd} N \arrow[r, "g"] \arrow[d, "f"'] & \mathcal{P}_1 \times_M \mathcal{P}_1 \arrow[d, "s\times t"] \\ (\mathcal{P}_0 \times \Gamma_0) \times (\mathcal{P}_0 \times \Gamma_0) \arrow[r, "\tau\times\tau"'] & (\mathcal{P}_0 \times_M \mathcal{P}_0) \times (\mathcal{P}_0 \times_M \mathcal{P}_0) \end{tikzcd}$$

is commutative. For a fixed point $n \in N$ we put

$$((p_1, g_1), (p_2, g_2)) := g(n) \in (P \times \Gamma_0) \times_M (P \times \Gamma_0),$$
$$((y, g, \tilde{g}), (y', g', \tilde{g}')) := f(n) \in (Y \times \Gamma_0 \times \Gamma_0) \times (Y \times \Gamma_0 \times \Gamma_0).$$

The commutativity of the diagram implies $\chi(p_1) = \chi(p_2) = (y', y)$, so that there exists $\gamma' \in \Gamma_1$ with $p_2 = p_1 \circ \gamma'$. We define $\gamma := \mathrm{id}_{g_1^{-1}} \cdot \gamma' \cdot \mathrm{id}_{\alpha(p_2)^{-1}g_2}$, which yields a morphism $\gamma \in \Gamma_1$ satisfying $\tau(p_1, g_1, \gamma) = (p_1, g_1, p_2, g_2) = g(n)$. On the other hand, we check that

$$(s(p_1, g_1, \gamma), t(p_1, g_1, \gamma)) = \big(\pi_2(p_1), g_1, s(\gamma), \pi_1(p_1), \alpha(p_1)^{-1} g_1, t(\gamma)\big) = f(n),$$

using that $s(\gamma) = g_1^{-1} g_2$ and $t(\gamma) = g_1^{-1} \alpha(p_1)\alpha(p_2)^{-1} g_2$. Summarizing, we have defined a smooth map

$$\sigma : N \to \mathcal{P}_1 \times \Gamma_1 : n \mapsto (p_1, g_1, \gamma)$$

such that $\tau \circ \sigma = g$ and $(s \times t) \circ \sigma = f$. Now let $\sigma' : N \to \mathcal{P}_1 \times \Gamma_1$ be another such map, and let $\sigma'(n) =: (p_1', g_1', \gamma')$. The condition that $\tau(\sigma(n)) = g(n) = \tau(\sigma'(n))$ shows immediately that $p_1 = p_1'$ and $g_1 = g_1'$, and then that $p_1 \circ \gamma = p_1 \circ \gamma'$. But since the Γ-action on P is principal, we have $\gamma = \gamma'$. This shows $\sigma = \sigma'$. Summarizing, $\mathcal{P}_1 \times \Gamma_1$ is a pullback. □

Example 7.2.4. Suppose $\Gamma = \mathcal{B}\mathrm{U}(1)$ (see Example 2.1.1(b)) and suppose $\mathcal{G}$ is a Γ-bundle gerbe over M, also known as a U(1)-bundle gerbe (see Example 5.1.7). Then, the associated principal $\mathcal{B}\mathrm{U}(1)$-2-bundle $\mathscr{R}_{\mathcal{G}}$ has the groupoid $\mathcal{P}$ with $\mathcal{P}_0 = Y$ and $\mathcal{P}_1 = P$, source and target maps $s = \pi_2 \circ \chi$ and $t = \pi_1 \circ \chi$, and composition $p_2 \circ p_1 = \mu(p_1, p_2)$. The action of $\mathcal{B}\mathrm{U}(1)$ on $\mathcal{P}$ is trivial on the level of objects and the given U(1)-action on P on the level of morphisms. The same applies for general abelian Lie groups A instead of U(1).

Construction of the 1*-isomorphism* $\mathcal{A}_{\mathcal{G}} : \mathcal{G} \to \mathscr{E}_M(\mathcal{R}_{\mathcal{G}})$. The Γ-bundle gerbe $\mathscr{E}_M(\mathcal{R}_{\mathcal{G}})$ has the surjective submersion $\tilde{Y} := Y \times \Gamma_0$ with $\tilde{\pi}(y, g) := \pi(y)$. The total space of its Γ-bundle $\tilde{P}$ is $\tilde{P} := P \times \Gamma_0 \times \Gamma_0$; it has the anchor $\alpha(p, g, h) = h$, the bundle projection

$$\tilde{\chi} : \tilde{P} \to \tilde{Y}^{[2]} : (p, g, h) \mapsto \big((\pi_1(\chi(p)), \alpha(p)^{-1}g), (\pi_2(\chi(p)), gh^{-1})\big),$$

the Γ-action is

$$\begin{aligned}
(p, g, h) \circ \gamma &\overset{(7.1\text{-}1)}{=} \big((p, g) \circ R((t_{\pi_2(\chi(p))}, gh^{-1}), \gamma), s(\gamma)\big)\\
&\overset{(7.2.2)}{=} \big((p, g) \circ (t_{\pi_2(\chi(p))} \circ (\mathrm{id}_{gh^{-1}} \cdot \gamma \cdot \mathrm{id}_{g^{-1}}), gh^{-1}s(\gamma)), s(\gamma)\big)\\
&\overset{(7.2.1)}{=} \big(\mu(t_{\pi_2(\chi(p))} \circ (\mathrm{id}_{gh^{-1}} \cdot \gamma \cdot \mathrm{id}_{g^{-1}}), p), gh^{-1}s(\gamma), s(\gamma)\big)\\
&\overset{(2.4\text{-}3)}{=} \big(p \circ (\mathrm{id}_{gh^{-1}} \cdot \gamma \cdot \mathrm{id}_{g^{-1}\alpha(p)}), gh^{-1}s(\gamma), s(\gamma)\big),
\end{aligned}$$

and its bundle gerbe product $\tilde{\mu}$ is given by

$$\begin{aligned}
\tilde{\mu}((p_{23}, g_{23}, h_{23}), (p_{12}, g_{12}, h_{12})) &\overset{(7.1\text{-}3)}{=} \big((p_{12}, g_{12}) \circ R((p_{23}, g_{23}), \mathrm{id}_{h_{12}}), h_{23}h_{12}\big)\\
&\overset{(7.2.2)}{=} ((p_{12}, g_{12}) \circ (p_{23}, g_{23}h_{12}), h_{23}h_{12})\\
&\overset{(7.2.1)}{=} (\mu(p_{23}, p_{12}), g_{23}h_{12}, h_{23}h_{12}).
\end{aligned}$$

In order to compare the bundle gerbes $\mathcal{G}$ and $\mathscr{E}_M(\mathcal{R}_{\mathcal{G}})$ we consider the smooth maps $\sigma : Y \to Y \times \Gamma_0$ and $\tilde{\sigma} : P \to \tilde{P}$ that are defined by $\sigma(y) := (y, 1)$ and $\tilde{\sigma}(p) := (p, \alpha(p), \alpha(p))$.

Lemma 7.2.5. *The map $\tilde{\sigma}$ defines an isomorphism $\tilde{\sigma} : P \to (\sigma \times \sigma)^* \tilde{P}$ of Γ-bundles over $Y^{[2]}$. Moreover, the diagram*

$$\begin{tikzcd}
\pi_{23}^* P \otimes \pi_{12}^* P \arrow[r, "\tilde{\sigma}\otimes\tilde{\sigma}"] \arrow[d, "\mu"'] & \tilde{\pi}_{23}^* \tilde{P} \otimes \tilde{\pi}_{12}^* \tilde{P} \arrow[d, "\tilde{\mu}"] \\
\pi_{13}^* P \arrow[r, "\tilde{\sigma}"'] & \tilde{\pi}_{13}^* \tilde{P}
\end{tikzcd}$$

is commutative.

Proof. For the first part it suffices to prove that $\tilde{\sigma}$ is Γ-equivariant, preserves the anchors, and that the diagram

$$\begin{tikzcd} P \arrow[r, "\tilde{\sigma}"] \arrow[d, "\chi"'] & \tilde{P} \arrow[d, "\tilde{\chi}"] \\ Y^{[2]} \arrow[r, "\sigma\times\sigma"'] & \tilde{Y}^{[2]} \end{tikzcd}$$

is commutative. Indeed, the commutativity of the diagram is obvious, and also that the anchors are preserved. For the Γ-equivariance, we have

$$\tilde{\sigma}(p\circ\gamma) = (p\circ\gamma, s(\gamma), s(\gamma)) = (p, \alpha(p), \alpha(p))\circ\gamma = \tilde{\sigma}(p)\circ\gamma.$$

Finally, we calculate

$$\begin{aligned}\tilde{\mu}\big((p_{23}, \alpha(p_{23}), \alpha(p_{23})), (p_{12}, \alpha(p_{12}), \alpha(p_{12}))\big) \\ &= \big(\mu(p_{23}, p_{12}), \alpha(p_{23})\alpha(p_{12}), \alpha(p_{23})\alpha(p_{12})\big) \\ &= \big(\mu(p_{23}, p_{12}), \alpha(\mu(p_{23}, p_{12})), \alpha(\mu(p_{23}, p_{12}))\big)\end{aligned}$$

which shows the commutativity of the diagram. □

Via Lemma 5.2.7 the bundle morphism $\tilde{\sigma}$ defines the required 1-morphism $\mathcal{A}_{\mathcal{G}}$, and Lemma 5.2.3 guarantees that $\mathcal{A}_{\mathcal{G}}$ is a 1-*iso*morphism.

Construction of the 1*-morphism* $\mathcal{R}_{\mathcal{A}} : \mathcal{P} \to \mathcal{P}'$. Let $\mathcal{A} : \mathscr{E}_M(\mathcal{P}) \to \mathscr{E}_M(\mathcal{P}')$ be a 1-morphism between Γ-bundle gerbes obtained from principal Γ-2-bundles $\mathcal{P}$ and $\mathcal{P}'$ over M. By Lemma 5.2.8 we can assume that $\mathcal{A}$ consists of a principal Γ-bundle $\chi : Q \to Z$ with $Z = \mathcal{P}_0 \times_M \mathcal{P}'_0$, and some isomorphism β over $Z^{[2]}$. For preparation, we consider the fiber products $Z_r := \mathcal{P}_0 \times_M \mathcal{P}_0'^{[2]}$ and $Z_l := \mathcal{P}_0^{[2]} \times_M \mathcal{P}'_0$ with the obvious embeddings $\iota_l : Z_l \to Z$ and $\iota_r : Z_r \to Z$ obtained by doubling elements. Together with the trivialization of Corollary 5.2.6, the pullbacks of β along ι_l and ι_r yield bundle morphisms

$$\beta_l := \iota_l^*\beta : \mathrm{pr}_{13}^* Q \to \mathrm{pr}_{23}^* Q \otimes \mathrm{pr}_{12}^* P$$

and

$$\beta_r := \iota_r^*\beta : \mathrm{pr}_{23}^* P' \otimes \mathrm{pr}_{12}^* Q \to \mathrm{pr}_{13}^* Q,$$

where $P := \mathcal{P}_1 \times \Gamma_0$ and $P' := \mathcal{P}' \times \Gamma_0$ are the principal Γ-bundles of the Γ-bundle gerbes $\mathscr{E}_M(\mathcal{P})$ and $\mathscr{E}_M(\mathcal{P}')$, respectively.

Lemma 7.2.6. *The bundle morphisms β_l and β_r have the following properties*:

(i) *They commute with each other in these sense that the diagram*

$$\begin{tikzcd} P'_{p'_1,p'_2}\otimes Q_{p_1,p'_1} \arrow[rr, "\beta_r"] \arrow[d, "\mathrm{id}\otimes\beta_l"'] \arrow[drr, "\beta"] & & Q_{p_1,p'_2} \arrow[d, "\beta_l"] \\ P'_{p'_1,p'_2}\otimes Q_{p_2,p'_1}\otimes P_{p_1,p_2} \arrow[rr, "\beta_r\otimes\mathrm{id}"'] & & Q_{p_2,p'_2}\otimes P_{p_1,p_2} \end{tikzcd}$$

is commutative for all $((p_1, p'_1), (p_2, p'_2)) \in Z^{[2]}$.

(ii) β_l *is compatible with the bundle gerbe product* μ *in the sense that*

$$\beta_l|_{p_1,p_3,p'} = (\mathrm{id}\otimes\mu_{p_1,p_2,p_3})\circ(\beta_l|_{p_2,p_3,p'}\otimes\mathrm{id})\circ\beta_l|_{p_1,p_2,p'}$$

for all $(p_1, p_2, p_3, p') \in \mathcal{P}_0^{[3]}\times\mathcal{P}'_0$.

(iii) β_r *is compatible with the bundle gerbe product* μ' *in the sense that*

$$\beta_r|_{p,p'_1,p'_3}\circ(\mu'_{p'_1,p'_2,p'_3}\otimes\mathrm{id}) = \beta_r|_{p,p'_2,p'_3}\circ(\mathrm{id}\otimes\beta_r|_{p,p'_1,p'_2})$$

for all $(p, p'_1, p'_2, p'_3) \in \mathcal{P}_0\times\mathcal{P}_0'^{[3]}$.

Proof. The identities (ii) and (iii) follow by restricting the commutative diagram (5-1) to the submanifolds $\mathcal{P}_0^{[3]}\times\mathcal{P}'_0$ and $\mathcal{P}_0\times\mathcal{P}_0'^{[3]}$ of $Z^{[3]}$, respectively. Similarly, the commutativity of the two triangular subdiagrams in (i) follows by restricting (5-1) along appropriate embeddings $Z^{[2]}\to Z^{[3]}$. □

Now we are in position to define the anafunctor $\mathcal{R}_{\mathcal{A}}$. First, we consider the left action

$$\beta_0 : \Gamma_0\times Q\to Q : (g, q)\mapsto\beta_r((\mathrm{id}, g), q)$$

that satisfies $\alpha(\beta_0(g, q)) = g\alpha(q)$. The action β_0 is properly discontinuous and free because β_r is a bundle isomorphism. The quotient $F := Q/\Gamma_0$ is the total space of the anafunctor $\mathcal{R}_{\mathcal{A}}$ we want to construct. Left and right anchors of an element $q\in F$ with $\chi(q) = (p, p')$ are given by

$$\alpha_l(q) := p \quad\text{and}\quad \alpha_r(q) := R(p', \alpha(q)).$$

The actions are defined by

$$\rho_l(\rho, q) := \beta_l^{-1}(q, (\rho, 1)) \quad\text{and}\quad \rho_r(q, \rho') := \beta_r\big((R(\rho', \mathrm{id}_{\alpha(q)^{-1}}), 1), q\big).$$

The left action is invariant under the action β_0 because of Lemma 7.2.6(i). For the right action, invariance follows from Lemma 7.2.6(ii) and the identity

$$\mu'\big((R(\rho', \mathrm{id}_{\alpha(q)^{-1}g^{-1}}), 1), (\mathrm{id}, g)\big) \overset{(7.1\text{-}3)}{=} \mu'\big((\mathrm{id}, g), (R(\rho', \mathrm{id}_{\alpha(q)^{-1}}), 1)\big).$$

Lemma 7.2.7. *The above formulas define an anafunctor $F : \mathcal{P} \to \mathcal{P}'$.*

Proof. The compatibility between anchors and actions is easy to check. The axiom for the actions ρ_l and ρ_r follows from parts (ii) and (iii) of Lemma 7.2.6. Lemma 7.2.6(i) shows that the actions commute. It remains to prove that $\alpha_l : F \to \mathcal{P}_0$ is a principal $\mathcal{P}'$-bundle. Since α_l is a composition of surjective submersions, we only have to show that the map

$$\tau : F\,{}_{\alpha_r}\!\times_t \mathcal{P}' \to F\,{}_{\alpha_l}\!\times_{\alpha_l} F : (q, \rho') \mapsto (q, \rho_r(q, \rho'))$$

is a diffeomorphism. We construct an inverse map τ^{-1} as follows. For (q_1, q_2) with $\chi(q_1) = (p, p')$ and $\chi(q_2) = (p, \tilde{p}')$, choose a representative

$$((\tilde{\rho}', g'), \tilde{q}) := \beta_r|^{-1}_{p,p',\tilde{p}'}(q_2).$$

Such choices can be made locally in a smooth way, and the result will not depend on them. We have $\chi(\tilde{q}) = (p, p')$ that there exists a unique $\gamma \in \Gamma_1$ such that $q_1 = \tilde{q} \circ \gamma$. Now we put

$$\tau^{-1}(q_1, q_2) := (q_1, R(\tilde{\rho}', \gamma^{-1})).$$

The calculation of $\tau^{-1} \circ \tau$ is straightforward. For the calculation of $(\tau \circ \tau^{-1})(q_1, q_2)$ we have to compute in the second component

$$\begin{aligned}
&\beta_r\big((R(\tilde{\rho}', \gamma^{-1} \cdot \mathrm{id}_{\alpha(q_1)^{-1}}), 1), q_1\big) \\
&\quad = \beta_r\big((R(\tilde{\rho}', \gamma^{-1} \cdot \mathrm{id}_{\alpha(q_1)^{-1}}), 1) \circ (\gamma \cdot \mathrm{id}_{\alpha(\tilde{q})^{-1}}), \tilde{q}\big) = \beta_r\big((\tilde{\rho}', \alpha(q_1)\alpha(\tilde{q})^{-1}), \tilde{q}\big) \\
&\quad = \beta_0\big(\alpha(q_1)\alpha(\tilde{q})^{-1} g'^{-1}, \beta_r((\tilde{\rho}', g'), \tilde{q})\big) = \beta_0(\alpha(q_1)\alpha(\tilde{q})^{-1} g'^{-1}, q_2),
\end{aligned}$$

and this is equivalent to q_2. $\square$

In order to promote the anafunctor F to a 1-morphism between principal 2-bundles, we have to do two things: we have to check that F commutes with the projections of the bundle $\mathcal{P}_1$ and $\mathcal{P}_2$, and we have to construct a Γ-equivariant structure on F. For the first point we use Remark 6.1.6(b), whose criterion $\pi \circ \alpha_l = \pi \circ \alpha_r$ is clearly satisfied. For the second point we provide a smooth action $\rho : F \times \Gamma_1 \to F$ in the sense of Definition A.1 and use Lemma A.2, which provides a construction of a Γ-equivariant structure. The action is defined by

$$\rho(q, \gamma) := \beta_l^{-1}\big(q \circ (\mathrm{id}_{\alpha(q)} \cdot \gamma \cdot \mathrm{id}_{t(\gamma)^{-1}}), (\mathrm{id}_{R(\alpha_l(q), t(\gamma))}, t(\gamma))\big). \tag{7.2.3}$$

Lemma 7.2.8. *This defines a smooth action of Γ_1 on F in the sense of Definition A.1.*

Proof. Smoothness is clear from the definition. The identity

$$\rho(\rho(q, \gamma_1), \gamma_2) = \beta_l^{-1}\big(q \circ (\mathrm{id}_{\alpha(q)} \cdot \gamma_1 \cdot \gamma_2 \cdot \mathrm{id}_{t(\gamma_2)^{-1} t(\gamma_1)^{-1}}), (\mathrm{id}, t(\gamma_1 \cdot \gamma_2))\big) = \rho(q, \gamma_1 \cdot \gamma_2)$$

follows from the definition and the two identities

$$\alpha(\rho(q,\gamma)) = \alpha(q)s(\gamma) \tag{7.2.4}$$
$$\text{and} \quad (\gamma_1 \cdot \mathrm{id}_{t(\gamma_1)^{-1}}) \cdot (\mathrm{id}_{s(\gamma_1)} \cdot \gamma_2 \cdot \mathrm{id}_{t(\gamma_2)^{-1}t(\gamma_1)^{-1}}) = \gamma_1 \cdot \gamma_2 \cdot \mathrm{id}_{t(\gamma_2)^{-1}t(\gamma_1)^{-1}}.$$

The latter can easily be verified upon substituting a crossed module for Γ. Checking condition (i) of Definition A.1 just uses the definitions. We check condition (ii) in two steps. First we prove the identity

$$\rho(\rho_l(\rho,q),\gamma_l \circ \gamma) = \rho_l(R(\rho,\gamma_l),\rho(q,\gamma)).$$

The main ingredient is the decomposition

$$\mathrm{id}_{\alpha(q)} \cdot (\gamma_l \circ \gamma) \cdot \mathrm{id}_{t(\gamma_l)^{-1}} = (\mathrm{id}_{\alpha(q)} \cdot \gamma \cdot \mathrm{id}_{t(\gamma)^{-1}}) \circ (\mathrm{id}_{\alpha(q)s(\gamma)t(\gamma)^{-1}} \cdot \gamma_l \cdot \mathrm{id}_{t(\gamma_l)^{-1}}) \tag{7.2.5}$$

that can, e.g., be verified in the crossed module language. Now we compute

$$\begin{aligned}
\rho(\rho_l(\rho,q),\gamma_l \circ \gamma) &= \beta_l^{-1}\big(q \circ (\mathrm{id}_{\alpha(q)} \cdot (\gamma_l \circ \gamma) \cdot \mathrm{id}_{t(\gamma_l)^{-1}}), (R(\rho,t(\gamma_l)),t(\gamma_l))\big) \\
&\overset{(7.2.5)}{=} \beta_l^{-1}\big(q \circ (\mathrm{id}_{\alpha(q)} \cdot \gamma \cdot \mathrm{id}_{t(\gamma)^{-1}}), (R(\rho,\gamma_l),t(\gamma_l))\big) \\
&= \rho_l(R(\rho,\gamma_l),\rho(q,\gamma)).
\end{aligned}$$

The second step is to show the identity

$$\rho(\rho_r(q,\rho'),\gamma \circ \gamma_r) = \rho_r(\rho(q,\gamma),R(\rho',\gamma_r)).$$

Here we use the decomposition

$$\mathrm{id}_{\alpha(q)} \cdot (\gamma \circ \gamma_r) \cdot \mathrm{id}_{t(\gamma)^{-1}} = (\mathrm{id}_{\alpha(q)} \cdot \gamma \cdot \mathrm{id}_{t(\gamma)^{-1}}) \circ (\mathrm{id}_{\alpha(q)} \cdot \gamma_r \cdot \mathrm{id}_{t(\gamma)^{-1}}). \tag{7.2.6}$$

Then we compute

$$\begin{aligned}
&\rho(\rho_r(q,\rho'),\gamma \circ \gamma_r) \\
&\quad = \beta_l^{-1}\big(\beta_r((R(\rho',\mathrm{id}_{\alpha(q)^{-1}}),1), q \circ (\mathrm{id}_{\alpha(q)} \cdot (\gamma \circ \gamma_r) \cdot \mathrm{id}_{t(\gamma)^{-1}})), (\mathrm{id},t(\gamma))\big) \\
&\quad \overset{(7.2.6)}{=} \beta_l^{-1}\big(\beta_r((R(\rho',\gamma_r \cdot \mathrm{id}_{s(\gamma)^{-1}\alpha(q)^{-1}}),1), \\
&\qquad\qquad \beta_0(\alpha(q)s(\gamma_r)s(\gamma)^{-1}\alpha(q)^{-1}, q \circ (\mathrm{id}_{\alpha(q)} \cdot \gamma \cdot \mathrm{id}_{t(\gamma)^{-1}}))), (\mathrm{id},t(\gamma))\big) \\
&\quad \overset{(7.2.4)}{=} \beta_l^{-1}\big(\beta_r((R(\rho',\gamma_r \cdot \mathrm{id}_{\alpha(\rho(q,\gamma))^{-1}})), q \circ (\mathrm{id}_{\alpha(q)} \cdot \gamma \cdot \mathrm{id}_{t(\gamma)^{-1}})), (\mathrm{id},t(\gamma))\big) \\
&\quad = \rho_r(\rho(q,\gamma),R(\rho',\gamma_r)),
\end{aligned}$$

where we have employed the equivalence relation on F that was generated by the action of β_0. □

Construction of a 2-isomorphism $\eta_{\mathcal{A}} : \mathcal{A} \Rightarrow \mathscr{E}_M(\mathscr{R}_{\mathcal{A}})$. We may again assume that the common refinement of $\mathcal{A}$ is the fiber product $\mathcal{P}_0 \times_M \mathcal{P}'_0$; otherwise, the proof of Lemma 5.2.8 provides a 2-isomorphism between $\mathcal{A}$ and one of these. Now, $\mathcal{A}$ and $\mathscr{E}_M(\mathscr{R}_{\mathcal{A}})$ have the same common refinement, and $\eta_{\mathcal{A}}$ is given by the map

$$\eta : Q \to F \times \Gamma_0 : q \mapsto (q, \alpha(q)).$$

This is obviously smooth and respects the projections to the base: if $\chi(q) = (p, p')$, then

$$\chi(q, \alpha(q)) \overset{(7.1.6)}{=} \big(\alpha_l(q), R(\alpha_r(q), \alpha(q)^{-1})\big) = (p, p').$$

Further, it respects the Γ-actions:

$$\eta(q\circ\gamma) = (q\circ\gamma, s(\gamma)) = \beta_l^{-1}(q\circ\gamma, (\mathrm{id}, 1)) \overset{(7.2.3)}{=} (\rho(q, \mathrm{id}_{\alpha(q)^{-1}}\cdot\gamma), s(\gamma)) \overset{(7.1.6)}{=} \eta(q)\circ\gamma,$$

so that η is a bundle morphism. It remains to verify the commutativity of the compatibility diagram (5-4). Let $((\rho', g'), q') \in P' \otimes \zeta_1^* Q$, and let $(q, (\rho, g)) \in \zeta_2^* Q \otimes P$ be a representative for $\beta((\rho', g'), q')$. In particular, we have $\alpha(q)g = g'\alpha(q')$, since β_r is anchor-preserving. Then, we get clockwise

$$(\eta \otimes \mathrm{id})\big(\beta((\rho', g'), q')\big) = \big((q, \alpha(q)), (\rho, g)\big). \tag{7.2.7}$$

Counterclockwise, we have to use the isomorphism of Lemma 7.1.5 that we call $\tilde{\beta}$ here. Then,

$$\begin{aligned}\tilde{\beta}\big((\mathrm{id} \otimes \eta)((\rho', g'), q')\big) &= \tilde{\beta}\big((\rho', g'), (q', \alpha(q'))\big)\\ &= \big((\tilde{q}, g'\alpha(q')g^{-1}), (\rho, g)\big)\end{aligned} \tag{7.2.8}$$

where the choices $(\tilde{\rho}, h)$ we have to make for the definition of $\tilde{\beta}$ are here (ρ, g^{-1}), and $\tilde{q}$ is defined in (7.1.7), which gives here

$$\tilde{q} = \beta_l^{-1}\big(\beta_r((\rho', 1), q'), (R(\rho^{-1}, \mathrm{id}_{g^{-1}}), g^{-1})\big).$$

Comparing (7.2.7) and (7.2.8) it remains to prove $q = \tilde{q}$ in F. As F was the quotient of Q by the action β_0, it suffices to have

$$\begin{aligned}\beta_0(g', \tilde{q}) &\overset{\text{(i)}}{=} \beta_l^{-1}\big(\beta_r((\mathrm{id}, g'), \beta_r((\rho', 1), q')), (R(\rho^{-1}, \mathrm{id}_{g^{-1}}), g^{-1})\big)\\ &\overset{\text{(iii)}}{=} \beta_l^{-1}\big(\beta_r((\rho', g'), q'), (R(\rho^{-1}, \mathrm{id}_{g^{-1}}), g^{-1})\big)\\ &= \beta_l^{-1}\big(\beta_l^{-1}(q, (\rho, g)), (R(\rho^{-1}, \mathrm{id}_{g^{-1}}), g^{-1})\big)\\ &\overset{\text{(ii)}}{=} \beta_l^{-1}(q, (\mathrm{id}, 1)) = q.\end{aligned}$$

This finishes the construction of the 2-isomorphism $\eta_{\mathcal{A}}$.

Appendix A. Equivariant anafunctors and group actions

In this section we are concerned with a Lie 2-group Γ and Lie groupoids $\mathcal{X}$ and $\mathcal{Y}$ with actions $R_1 : \mathcal{X} \times \Gamma \to \mathcal{X}$ and $R_2 : \mathcal{Y} \times \Gamma \to \mathcal{Y}$.

Definition A.1. An *action* of the 2-group Γ on an anafunctor $F : \mathcal{X} \to \mathcal{Y}$ is an ordinary smooth action $\rho : F \times \Gamma_1 \to F$ of the group Γ_1 on the total space F that

(i) preserves the anchors in the sense that the diagrams

$$\begin{tikzcd} F \times \Gamma_1 \arrow[r, "\rho"] \arrow[d, "\alpha_l \times t"'] & F \arrow[d, "\alpha_l"] \\ \mathcal{X}_0 \times \Gamma_0 \arrow[r, "R_1"'] & \mathcal{X}_0 \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} F \times \Gamma_1 \arrow[r, "\rho"] \arrow[d, "\alpha_r \times s"'] & F \arrow[d, "\alpha_r"] \\ \mathcal{Y}_0 \times \Gamma_0 \arrow[r, "R_2"'] & \mathcal{Y}_0 \end{tikzcd}$$

are commutative;

(ii) is compatible with the Γ-actions in the sense that the identity

$$\rho(\chi \circ f \circ \eta, \gamma_l \circ \gamma \circ \gamma_r) = R_1(\chi, \gamma_l) \circ \rho(f, \gamma) \circ R_2(\eta, \gamma_r)$$

holds for all appropriately composable $\chi \in \mathcal{X}_1$, $\eta \in \mathcal{Y}_1$, $f \in F$, and $\gamma_l, \gamma, \gamma_r \in \Gamma_1$.

If $F_1, F_2 : \mathcal{X} \to \mathcal{Y}$ are anafunctors with Γ-action, a transformation $\eta : F_1 \Rightarrow F_2$ is called Γ-equivariant if the map $\eta : F_1 \to F_2$ between total spaces is Γ_1-equivariant in the ordinary sense.

Anafunctors $\mathcal{X} \to \mathcal{Y}$ with Γ-actions together with Γ-equivariant transformations form a groupoid $\mathcal{A}\mathrm{na}^{\infty}_{\Gamma}(\mathcal{X}, \mathcal{Y})$. On the other hand, there is another groupoid $\Gamma\text{-}\mathcal{A}\mathrm{na}^{\infty}(\mathcal{X}, \mathcal{Y})$ consisting of Γ-equivariant anafunctors (Definition 6.1.3) and Γ-equivariant transformations (Definition 6.1.4).

Lemma A.2. *The categories $\mathcal{A}\mathrm{na}^{\infty}_{\Gamma}(\mathcal{X}, \mathcal{Y})$ and $\Gamma\text{-}\mathcal{A}\mathrm{na}^{\infty}(\mathcal{X}, \mathcal{Y})$ are canonically isomorphic.*

Proof. We construct a functor

$$\mathcal{E} : \mathcal{A}\mathrm{na}^{\infty}_{\Gamma}(\mathcal{X}, \mathcal{Y}) \to \Gamma\text{-}\mathcal{A}\mathrm{na}^{\infty}(\mathcal{X}, \mathcal{Y}). \tag{A-1}$$

Let $F : \mathcal{X} \to \mathcal{Y}$ be an anafunctor with Γ-action ρ. We shall define a transformation

$$\lambda_\rho : F \circ R_1 \Rightarrow R_2 \circ (F \times \mathrm{id}).$$

First of all, the composite

$$\mathcal{X} \times \Gamma \xrightarrow{R_1} \mathcal{X} \xrightarrow{F} \mathcal{Y}$$

is given by the total space $(\mathcal{X}_0 \times \Gamma_0)\ {}_{R_1}{\times}_{\alpha_l} F$, left and right anchors send an element (x, g, f) to (x, g) and $\alpha_r(f)$, respectively, and the actions are

$(\chi, \gamma) \circ (x, g, f) = (t(\chi), t(\gamma), R_1(\chi, \gamma) \circ f)$ and $(x, g, f) \circ \eta = (x, g, f \circ \eta)$.

On the other hand, the composite

$$\mathcal{X} \times \Gamma \xrightarrow{F \times \mathrm{id}} \mathcal{Y} \times \Gamma \xrightarrow{R_2} \mathcal{Y}$$

is given by the total space $((F \times \Gamma_1) \,_{R_2 \circ (\alpha_r \times s)}\times_t \mathcal{Y}_1) / \sim$ with the equivalence relation

$$(f \circ \eta', \gamma \circ \gamma', \eta) \sim (f, \gamma, R_2(\eta', \gamma') \circ \eta).$$

The left and right anchors send an element (f, γ, η) to $(\alpha_l(f), t(\gamma))$ and $s(\eta)$, respectively, and the actions are

$$(\chi, \gamma') \circ (f, \gamma, \eta) = (\chi \circ f, \gamma' \circ \gamma, \eta) \quad \text{and} \quad (f, \gamma, \eta) \circ \eta' = (f, \gamma, \eta \circ \eta').$$

The inverse of the following map will define the transformation λ:

$$(F \times \Gamma_1) \,_{R_2 \circ (\alpha_r \times s)}\times_t \mathcal{Y}_1 \to (\mathcal{X}_0 \times \Gamma_0) \,_{R_1}\times_{\alpha_l} F : (f, \gamma, \eta) \mapsto \big(\alpha_l(f), t(\gamma), \rho(f, \gamma) \circ \eta\big).$$

Condition (i) ensures that this map ends in the correct fiber product, and condition (ii) ensures that it is well-defined under the equivalence relation $\sim$. The left anchors are automatically respected, and the right anchors require condition (i). Similarly, the left action is respected automatically, and the right actions due to condition (ii). The axiom for a transformation is satisfied because ρ is a group action. This defines the functor $\mathcal{E}$ on objects. On morphisms, it is straightforward to check that the conditions on both hand sides coincide; in particular, $\mathcal{E}$ is full and faithful.

In order to prove that the functor $\mathcal{E}$ is an isomorphism, we start with a given Γ-equivariant structure λ on the anafunctor F. Then, an action $\rho : F \times \Gamma_1 \to F$ is defined by

$$(f, \gamma) \mapsto \mathrm{pr}_3(\lambda^{-1}(f, \gamma, \mathrm{id}_{R_2(\alpha_r(f), s(\gamma))}))$$

with $\mathrm{pr}_3 : (\mathcal{X}_0 \times \Gamma_0) \,_{R_1}\times_{\alpha_l} F \to F$ the projection. The axiom for an action is satisfied due to the identity λ obeys. It is straightforward to verify conditions (i) and (ii) of Definition A.1. To close the proof it suffices to notice that the two procedures we have defined are (strictly) inverse to each other. □

We are also concerned with the composition of anafunctors with Γ-action. Suppose that $\mathcal{Z}$ is a third Lie groupoid with a Γ-action R_3, and $F : \mathcal{X} \to \mathcal{Y}$ and $G : \mathcal{Y} \to \mathcal{Z}$ are anafunctors with Γ-actions $\rho : F \times \Gamma_1 \to F$ and $\tau : G \times \Gamma_1 \to G$. Then, the composition $G \circ F$ is equipped with the Γ-action defined by

$$(F \times_{\mathcal{Y}_0} G) \times \Gamma_1 \to (F \times_{\mathcal{Y}_0} G) : ((f, g), \gamma) \mapsto (\rho(f, \gamma), \tau(g, \mathrm{id}_{s(\gamma)})). \tag{A-2}$$

We leave it to the reader to check:

Lemma A.3. *Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be Lie groupoids with Γ-actions R_1, R_2 and R_3.*

(a) *Let* $F : \mathcal{X} \to \mathcal{Y}$ *and* $G : \mathcal{Y} \to \mathcal{Z}$ *be* Γ*-equivariant anafunctors. If* Γ*-equivariant structures on* F *and* G *correspond to* Γ_1*-actions under the isomorphism of Lemma A.2, then the* Γ*-equivariant structure on the composite* $F \circ G$ *corresponds to the* Γ_1*-action defined above.*

(b) *The isomorphism of Lemma A.2 identifies the trivial* Γ*-equivariant structure on the identity anafunctor* $\mathrm{id} : \mathcal{X} \to \mathcal{X}$ *with the* Γ_1*-action* $R_1 : \mathcal{X}_1 \times \Gamma_1 \to \mathcal{X}_1$ *on its total space* $\mathcal{X}$.

Appendix B. Constructing equivalences between 2-stacks

Let $\mathcal{C}$ be a bicategory (we assume that associators and unifiers are *invertible* 2-morphisms). We fix the following terminology: a 1-*iso*morphism $f : X_1 \to X_2$ in $\mathcal{C}$ always includes the data of an inverse 1-morphism $\bar{f} : X_2 \to X_1$ and of 2-isomorphisms $i : \bar{f} \circ f \Rightarrow \mathrm{id}$ and $j : \mathrm{id} \Rightarrow f \circ \bar{f}$ satisfying the zigzag identities. Let $\mathcal{D}$ be another bicategory. A 2-functor $F : \mathcal{C} \to \mathcal{D}$ is assumed to have *invertible* compositors and unitors.

The following lemma is certainly “well-known”, although we have not been able to find a reference for exactly this statement.

Lemma B.1. *Let* $F : \mathcal{C} \to \mathcal{D}$ *be a* 2*-functor that is fully faithful on Hom-categories. Suppose one has chosen*:

(1) *for every object* $Y \in \mathcal{D}$ *an object* $G_Y \in \mathcal{C}$ *and a* 1*-isomorphism* $\xi_Y : Y \to F(G_Y)$;

(2) *for all objects* $X_1, X_2 \in \mathcal{C}$ *and all* 1*-morphisms* $g : F(X_1) \to F(X_2)$, *a* 1*-morphism* $G_g : X_1 \to X_2$ *in* $\mathcal{C}$ *together with a* 2*-isomorphism* $\eta_g : g \Rightarrow F(G_g)$.[1]

Then, there is a 2*-functor* $G : \mathcal{D} \to \mathcal{C}$ *and pseudonatural equivalences*

$$a : \mathrm{id}_{\mathcal{D}} \Rightarrow F \circ G \quad \textit{and} \quad b : G \circ F \Rightarrow \mathrm{id}_{\mathcal{C}}.$$

In particular, F is an equivalence of bicategories.

Proof. We recall our convention concerning 1-isomorphisms: the 1-isomorphisms ξ_Y include choices of inverse 1-morphisms $\bar{\xi}_Y$ together with 2-isomorphisms $i_Y : \bar{\xi}_Y \circ \xi_Y \Rightarrow \mathrm{id}$ and $j_Y : \mathrm{id} \Rightarrow \xi_Y \circ \bar{\xi}_Y$ satisfying the zigzag identities.

First we explicitly construct the 2-functor G. On objects, we put $G(Y) := G_Y$. We use the notation $\tilde{g} := (\xi_{Y_2} \circ g) \circ \bar{\xi}_{Y_1}$ for all 1-morphisms $g : Y_1 \to Y_2$ in $\mathcal{D}$, and define $G(g) = G_{\tilde{g}}$. If $g, g' : Y_1 \to Y_2$ are 1-morphisms, and $\psi : g \Rightarrow g'$ is a 2-morphism, we consider the 2-morphism $\tilde{\psi}$ defined by

$$F(G_{\tilde{g}}) \xRightarrow{\eta_{\tilde{g}}^{-1}} (\xi_{Y_2} \circ g) \circ \bar{\xi}_{Y_1} \xRightarrow{(\mathrm{id} \circ \psi) \circ \mathrm{id}} (\xi_{Y_2} \circ g') \circ \bar{\xi}_{Y_1} \xRightarrow{\eta_{\tilde{g}'}} F(G_{\tilde{g}'}).$$

[1] More accurately we should write $G_{X_1,X_2,g}$ and $\eta_{X_1,X_2,g}$, but we will suppress X_1 and X_2 in the notation.

Since F is fully faithful on 2-morphisms, we may choose the unique 2-morphism $G(\psi) : G(g) \Rightarrow G(g')$ such that $F(G(\psi)) = \tilde{\psi}$. In order to define the compositor of G we look at 1-morphisms $g_{12} : Y_1 \to Y_2$ and $g_{23} : Y_2 \to Y_3$. We consider the 2-morphism

$$\begin{aligned}
F(G(g_{23}) \circ G(g_{12})) &\xRightarrow{c^{-1}_{G(g_{12}),G(g_{23})}} F(G_{\tilde{g}_{23}}) \circ F(G_{\tilde{g}_{12}}) \\
&\Downarrow \eta^{-1}_{\tilde{g}_{23}} \circ \eta^{-1}_{\tilde{g}_{12}} \\
&((\xi_{Y_3} \circ g_{23}) \circ \bar{\xi}_{Y_2}) \circ ((\xi_{Y_2} \circ g_{12}) \circ \bar{\xi}_{Y_1}) \\
&\Downarrow a, i_{Y_2} \\
&(\xi_{Y_3} \circ (g_{23} \circ g_{12})) \circ \bar{\xi}_{Y_1} \xRightarrow[\eta_{\widetilde{g_{23} \circ g_{12}}}]{} F(G(g_{23} \circ g_{12}));
\end{aligned}$$

its unique preimage under the 2-functor F is the compositor

$$c_{g_{12},g_{23}} : G(g_{23}) \circ G(g_{12}) \Rightarrow G(g_{23} \circ g_{12}).$$

In order to define the unitor of G we consider an object $Y \in \mathcal{D}$ and look at the 2-morphism

$$F(G(\mathrm{id}_Y)) \xRightarrow{\eta^{-1}_{\widetilde{\mathrm{id}_Y}}} (\xi_Y \circ \mathrm{id}_Y) \circ \bar{\xi}_Y \xRightarrow{l_{\xi_Y}, j_Y^{-1}} \mathrm{id}_{F(G(Y))} \xRightarrow{u^{-1}_{G(Y)}} F(\mathrm{id}_{G(Y)}).$$

Its unique preimage under the 2-functor F is the unitor $u_Y : G(\mathrm{id}_Y) \Rightarrow \mathrm{id}_{G(Y)}$. The second step is to verify the axioms of a 2-functor. This is simple but extremely tedious and can only be left as an exercise. The third step is to construct the pseudonatural transformation

$$a : \mathrm{id}_{\mathcal{D}} \Rightarrow F \circ G.$$

Its component at an object Y in $\mathcal{D}$ is the 1-morphism $a(Y) := \xi_Y : Y \to F(G(Y))$. Its component at a 1-morphism $g : Y_1 \to Y_2$ is the 2-morphism $a(g)$ defined by

$$\begin{aligned}
a(Y_2) \circ g &=\!=\!=\!= \xi_{Y_2} \circ g \\
&\Downarrow \mathrm{id} \circ l^{-1}_{\xi_{Y_2} \circ g} \\
&(\xi_{Y_2} \circ g) \circ \mathrm{id} \\
&\Downarrow a, i^{-1}_{Y_2} \\
&((\xi_{Y_2} \circ g) \circ \bar{\xi}_{Y_1}) \circ \xi_{Y_1} \\
&\Downarrow \eta_{\tilde{g}} \circ \mathrm{id} \\
&F(G_{\tilde{g}}) \circ \xi_{Y_1} =\!=\!=\!= F(G(g)) \circ a(Y_1).
\end{aligned}$$

There are two axioms a pseudonatural transformation has to satisfy, and their proofs are again left as an exercise. It is easy to see that a is a pseudonatural *equivalence*, with an inverse transformation given by $\bar{a}(Y) := \bar{\xi}_Y$. The fourth and last step is to construct the pseudonatural transformation

$$b : G \circ F \Rightarrow \mathrm{id}_{\mathcal{C}}.$$

Its component at an object X is $b(X) := G_{\bar{\xi}_{F(X)}} : G(F(X)) \to X$. Its component at a 1-morphism $f : X_2 \to X_2$ is the 2-morphism

$$b(f) : b(X_2) \circ G(F(f)) \Rightarrow f \circ b(X_1)$$

given as the unique preimage under F of the 2-morphism

$$\begin{aligned}
F(b(X_2) \circ G(F(f))) \xRightarrow{c^{-1}} {} & F(b(X_2)) \circ F(G(F(f))) \\
& \Big\Downarrow{\scriptstyle \eta^{-1}_{\bar{\xi}_{F(X_2)}} \circ \eta^{-1}_{F(f)}} \\
& \bar{\xi}_{F(X_2)} \circ ((\xi_{F(X_2)} \circ F(f)) \circ \xi_{F(X_1)}) \\
& \Big\Downarrow{\scriptstyle a, i_{F(X_2)}, r} \\
& F(f) \circ \bar{\xi}_{F(X_1)} \\
& \Big\Downarrow{\scriptstyle \mathrm{id}_{F(f)} \circ \eta_{\bar{\xi}_{F(X_1)}}} \\
& F(f) \circ F(b(X_1)) \xRightarrow[c]{} F(f \circ b(X_1)).
\end{aligned}$$

The proofs of the axioms are again left for the reader, and again it is easy to see that b is a pseudonatural *equivalence* with an inverse transformation given by $\bar{b}(X) := G_{\xi_{F(X)}}$. □

As a consequence of Lemma B.1 we obtain the certainly well-known result:

Corollary B.2. *Let $F : \mathcal{C} \to \mathcal{D}$ be essentially surjective, and an equivalence on all Hom-categories. Then, F is an equivalence of bicategories.*

Since we work with 2-stacks over manifolds, we need the following stacky extension of Lemma B.1. For a pre-2-stack $\mathcal{C}$, we denote by $\mathcal{C}_M$ the 2-category it associates to a smooth manifold M, and by $\psi^* : \mathcal{C}_N \to \mathcal{C}_M$ the 2-functor it associates to a smooth map $\psi : M \to N$. The pseudonatural equivalences $\psi^* \circ \varphi^* \cong (\varphi \circ \psi)^*$ will be suppressed from the notation in the following. If $\mathcal{C}$ and $\mathcal{D}$ are pre-2-stacks, a 1-*morphism* $F : \mathcal{C} \to \mathcal{D}$ associates 2-functors $F_M : \mathcal{C}_M \to \mathcal{D}_M$ to a smooth manifold M, pseudonatural equivalences

$$F_\psi : \psi^* \circ F_N \to F_M \circ \psi^*$$

to smooth maps $\psi : M \to N$, and certain modifications $F_{\psi,\varphi}$ that control the relation between F_ψ and F_φ for composable maps ψ and φ.

Lemma B.3. *Suppose $\mathcal{C}$ and $\mathcal{D}$ are pre-2-stacks over smooth manifolds, and $F : \mathcal{C} \to \mathcal{D}$ is a 1-morphism. Suppose that for every smooth manifold M*

(1) *the assumptions of Lemma B.1 for the 2-functor F_M are satisfied, and*

(2) *the data (G_Y, ξ_Y) and (G_g, η_g) is chosen for all objects Y and 1-morphisms g in $\mathcal{D}_M$.*

Then, there is a 1-morphism $G : \mathcal{D} \to \mathcal{C}$ of pre-2-stacks together with 2-isomorphisms

$$a : F \circ G \Rightarrow \mathrm{id}_{\mathcal{D}} \quad \textit{and} \quad b : G \circ F \Rightarrow \mathrm{id}_{\mathcal{C}}$$

such that for every smooth manifold M the 2-functor G_M and the pseudonatural transformations a_M and b_M are the ones of Lemma B.1. In particular, F is an equivalence of pre-2-stacks.

For the proof one constructs the required pseudonatural equivalences G_ψ and the modifications $G_{\psi,\varphi}$ from the given ones, F_ψ and $F_{\psi,\varphi}$, respectively, in a similar way as explained in the proof of Lemma B.1. Since these constructions are straightforward to do but would consume many pages, and the statement of the lemma is not too surprising and certainly well-known to many people, we leave these constructions for the interested reader.

Acknowledgements

We thank Christoph Wockel for helpful discussions. We also thank the Erwin Schrödinger Institute in Vienna and the Instituto Superior Técnico in Lisbon for kind invitations.

References

[Aldrovandi and Noohi 2009] E. Aldrovandi and B. Noohi, "Butterflies, I: Morphisms of 2-group stacks", *Adv. Math.* **221**:3 (2009), 687–773. MR 2010c:18009 Zbl 1179.18007 arXiv 0808.3627

[Aschieri et al. 2005] P. Aschieri, L. Cantini, and B. Jurčo, "Nonabelian bundle gerbes, their differential geometry and gauge theory", *Comm. Math. Phys.* **254**:2 (2005), 367–400. MR 2005k:53022 Zbl 1092.53020

[Baas et al. 2012] N. A. Baas, M. Bökstedt, and T. A. Kro, "Two-categorical bundles and their classifying spaces", *J. K-Theory* **10**:2 (2012), 299–369. Zbl 06111751 arXiv 0612549

[Baez and Schreiber 2007] J. C. Baez and U. Schreiber, "Higher gauge theory", pp. 7–30 in *Categories in algebra, geometry and mathematical physics*, edited by A. Davydov, Contemp. Math. **431**, Amer. Math. Soc., Providence, RI, 2007. MR 2008k:53099 Zbl 1132.55007 arXiv math/0511710

[Baez and Stevenson 2009] J. C. Baez and D. Stevenson, "The classifying space of a topological 2-group", pp. 1–31 in *Algebraic topology*, edited by N. A. Baas, Abel Symp. **4**, Springer, Berlin, 2009. MR 2011h:55027 Zbl 1182.55015 arXiv 0801.3843

[Baez et al. 2007] J. C. Baez, D. Stevenson, A. S. Crans, and U. Schreiber, "From loop groups to 2-groups", *Homology Homotopy Appl.* **9**:2 (2007), 101–135. MR 2009c:22022 Zbl 1122.22003 arXiv math.QA/0504123

[Bartels 2006] T. K. Bartels, *Higher gauge theory: 2-bundles*, Ph.D. thesis, University of California, Riverside, CA, 2006, Available at http://tinyurl.com/bartels-phd. MR 2709030 arXiv math/0410328

[Bénabou 1967] J. Bénabou, "Introduction to bicategories", pp. 1–77 in *Reports of the Midwest Category Seminar* (Chicago, 1967), Lecture Notes in Math. **47**, Springer, Berlin, 1967. MR 36 #3841

[Bénabou 1973] J. Bénabou, *Les distributeurs: d'après le cours de questions spéciales de mathématique*, Seminaires de Math. Pure **33**, Institut de Mathématique pure et appliquée, Université Catholique de Louvain, Louvain-la-Neuve, 1973.

[Breen 1990] L. Breen, "Bitorseurs et cohomologie non abélienne", pp. 401–476 in *The Grothendieck Festschrift I*, edited by P. Cartier et al., Progr. Math. **86**, Birkhäuser, Boston, 1990. MR 92m:18019 Zbl 0743.14034

[Breen 1994] L. Breen, *On the classification of 2-gerbes and 2-stacks*, Astérisque **225**, Société de Mathématique de France, Paris, 1994. MR 95m:18006 Zbl 0818.18005

[Bunke 2011] U. Bunke, "String structures and trivialisations of a Pfaffian line bundle", *Comm. Math. Phys.* **307**:3 (2011), 675–712. MR 2842963 Zbl 1238.58022 arXiv 0909.0846

[Ginot and Stiénon 2008] G. Ginot and M. Stiénon, "G-Gerbes, principal 2-group bundles and characteristic classes", preprint, 2008. arXiv 0801.1238

[Giraud 1971] J. Giraud, *Cohomologie non abélienne*, Grundlehren der Math. Wiss. **179**, Springer, Berlin, 1971. MR 49 #8992 Zbl 0226.14011

[Heinloth 2005] J. Heinloth, "Notes on differentiable stacks", pp. 1–32 in *Seminars Winter Term* 2004/2005 (Göttingen, 2004–2005), edited by Y. Tschinkel, Universitätsverlag Göttingen, Göttingen, 2005. MR 2007c:19009 Zbl 1098.14501

[Johnstone 1977] P. T. Johnstone, *Topos theory*, London Math. Soc. Monogr. **10**, Academic Press, London, 1977. MR 57 #9791 Zbl 0368.18001

[Jurčo 2011] B. Jurčo, "Crossed module bundle gerbes: classification, string group and differential geometry", *Int. J. Geom. Methods Mod. Phys.* **8**:5 (2011), 1079–1095. MR 2838538 Zbl 1232.53027 arXiv math/0510078

[Laurent-Gengoux et al. 2009] C. Laurent-Gengoux, M. Stiénon, and P. Xu, "Non-abelian differentiable gerbes", *Adv. Math.* **220**:5 (2009), 1357–1427. MR 2010c:53039 Zbl 1177.22001 arXiv math/0511696

[Lerman 2010] E. Lerman, "Orbifolds as stacks?", *Enseign. Math.* (2) **56**:3-4 (2010), 315–363. MR 2012c:18010 Zbl 1221.14003 arXiv 0806.4160

[Metzler 2003] D. S. Metzler, "Topological and smooth stacks", preprint, 2003. arXiv math/0306176

[Milnor 1956] J. Milnor, "Construction of universal bundles, II", *Ann. of Math.* (2) **63** (1956), 430–436. MR 17,1120a Zbl 0071.17401

[Moerdijk and Mrčun 2003] I. Moerdijk and J. Mrčun, *Introduction to foliations and Lie groupoids*, Cambridge Stud. Adv. Math. **91**, Cambridge University Press, Cambridge, 2003. MR 2005c:58039 Zbl 1029.58012

[Müller and Wockel 2009] C. Müller and C. Wockel, "Equivalences of smooth and continuous principal bundles with infinite-dimensional structure group", *Adv. Geom.* **9**:4 (2009), 605–626. MR 2010k:22026 Zbl 1184.22011 arXiv math/0604142

[Murray 1996] M. K. Murray, "Bundle gerbes", *J. London Math. Soc.* (2) **54**:2 (1996), 403–416. MR 98a:55016 Zbl 0867.55019 arXiv dg-ga/9407015

[Murray and Stevenson 2000] M. K. Murray and D. Stevenson, "Bundle gerbes: stable isomorphism and local theory", *J. London Math. Soc.* (2) **62**:3 (2000), 925–937. MR 2001j:55019 Zbl 1019.55009 arXiv math/9908135

[Murray et al. 2012] M. K. Murray, D. M. Roberts, and D. S. Stevenson, "On the existence of bibundles", *Proc. Lond. Math. Soc.* **105**:6 (2012), 1290–1314. Zbl 06121966 arXiv 1102.4388

[Nikolaus and Schweigert 2011] T. Nikolaus and C. Schweigert, "Equivariance in higher geometry", *Adv. Math.* **226**:4 (2011), 3367–3408. MR 2012h:18004 Zbl 1219.22003 arXiv 1004.4558

[Onishchik and Vinberg 1988] A. L. Onishchik and È. B. Vinberg, "Foundations of Lie theory", pp. 5–101 in Группы Ли и алгебры Ли, edited by A. L. Onishchik and R. V. Gamkrelidze, Itogi Nauki Tekh., Ser. Sovrem. Probl. Mat., Fundam. Napravleniya **20**, Vsesoyuznyj Institut Nauchnoj i Tekhnicheskoj Informatsii, Moscow, 1988. In Russian. Translated on pp. 1–94 in *Lie groups and Lie algebras I*, edited by Onishchik, A. L., Encyclopaedia of Mathematical Sciences **20**, Springer, Berlin, 1993. MR 1306738 Zbl 0656.22002

[Pronk 1996] D. A. Pronk, "Etendues and stacks as bicategories of fractions", *Compositio Math.* **102**:3 (1996), 243–303. MR 97d:18011 Zbl 0871.18003

[Schommer-Pries 2011] C. J. Schommer-Pries, "Central extensions of smooth 2-groups and a finite-dimensional string 2-group", *Geom. Topol.* **15**:2 (2011), 609–676. MR 2012j:53024 Zbl 1216.22005 arXiv 0911.2483

[Schreiber and Waldorf 2008] U. Schreiber and K. Waldorf, "Connections on non-abelian gerbes and their holonomy", preprint, 2008. arXiv 0808.1923

[Schreiber et al. 2007] U. Schreiber, C. Schweigert, and K. Waldorf, "Unoriented WZW models and holonomy of bundle gerbes", *Comm. Math. Phys.* **274**:1 (2007), 31–64. MR 2009g:53074 Zbl 1148.53057 arXiv hep-th/0512283

[Segal 1968] G. Segal, "Classifying spaces and spectral sequences", *Inst. Hautes Études Sci. Publ. Math.* **34** (1968), 105–112. MR 38 #718 Zbl 0199.26404

[Stevenson 2000] D. Stevenson, *The geometry of bundle gerbes*, Ph.D. thesis, University of Adelaide, 2000. arXiv math/0004117

[Waldorf 2007] K. Waldorf, "More morphisms between bundle gerbes", *Theory Appl. Categ.* **18**:9 (2007), 240–273. MR 2008k:53101 Zbl 1166.55005 arXiv math.CT/0702652

[Wockel 2011] C. Wockel, "Principal 2-bundles and their gauge 2-groups", *Forum Math.* **23**:3 (2011), 565–610. MR 2012d:55023 Zbl 1232.55013 arXiv 0803.3692

Received April 23, 2012.

THOMAS NIKOLAUS
FACHBEREICH MATHEMATIK
BEREICH ALGEBRA UND ZAHLENTHEORIE
UNIVERSITÄT HAMBURG
BUNDESSTRASSE 55
D-20146 HAMBURG
GERMANY
thomas1.nikolaus@mathematik.uni-regensburg.de

KONRAD WALDORF
FAKULTÄT FÜR MATHEMATIK
UNIVERSITÄT REGENSBURG
UNIVERSITÄTSSTRASSE 31
D-93053 REGENSBURG
GERMANY
konrad.waldorf@mathematik.uni-regensburg.de

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.421

ON NONLINEAR NONHOMOGENEOUS RESONANT DIRICHLET EQUATIONS

NIKOLAOS S. PAPAGEORGIOU AND GEORGE SMYRLIS

We consider a $(p, 2)$-equation with a Carathéodory reaction $f(z, x)$ which is resonant at $\pm\infty$ and has constant sign, z-dependent zeros. Using variational methods, together with truncation and comparison techniques and Morse theory, we establish the existence of five nontrivial smooth solutions (four of constant sign and the fifth nodal). If the reaction $f(z, x)$ is C^1 in $x \in \mathbb{R}$, then we produce a second nodal solution for a total of six nontrivial smooth solutions.

1. Introduction

Let $\Omega \subseteq \mathbb{R}^N$ be a bounded domain with a C^2-boundary $\partial\Omega$. In this paper we study the nonlinear Dirichlet problem

$$-\Delta_p u(z) - \Delta u(z) = f(z, u(z)) \quad \text{in } \Omega, \quad u|_{\partial\Omega} = 0, \quad 2 < p. \tag{1}$$

Here Δ_p denotes the p-Laplacian differential operator defined by

$$\Delta_p u(z) = \operatorname{div}(\|Du(z)\|^{p-2} Du(z)) \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

Problem (1) is important in quantum physics in connection with Derrick's model [Derrick 1964] for the existence of solitons, which was investigated in more detail by Benci, D'Avenia, Fortunato, and Pisani [Benci et al. 2000]. Recently, such equations attracted the interest of people working on nonlinear partial differential equations and some existence and multiplicity results were proved in [Cingolani and Degiovanni 2005; Cingolani and Vannella 2003; Sun 2012]. All consider nonresonant equations. In contrast, in this work we deal with the resonant case. More precisely, we assume that, asymptotically at $\pm\infty$, we have resonance with respect to the first eigenvalue of $(-\Delta_p, W_0^{1,p}(\Omega))$. In problem (1) the reaction $f(z, x)$ is a Carathéodory function (i.e., for all $x \in \mathbb{R}$, $z \to f(z, x)$ is measurable, and, for a.a. $z \in \Omega$, $x \to f(z, x)$ is continuous) and has positive and negative zeros which in general depend on $z \in \Omega$. Problems driven by the p-Laplacian, and with a

MSC2010: 35J20, 35J60, 35J92, 58E05.
Keywords: resonant equations, tangency principle, strong comparison principle, constant sign and nodal solutions, Morse theory.

reaction that has zeros, were studied by Bartsch, Liu, and Weth [Bartsch et al. 2005] (they assume that the zeros are constant) and by Iturriaga, Massa, Sánchez, and Ubilla [Iturriaga et al. 2010] (they have variable zeros). In both works the reaction $f(z,\cdot)$ is $(p-1)$-superlinear.

Here, we prove the existence of at least five nontrivial smooth solutions and provide sign information for all of them (two are positive, two are negative and the fifth is nodal). Moreover, by strengthening the regularity of $f(z,\cdot)$ (namely, assuming that $f(z,\cdot)\in C^1(\mathbb{R})$), we produce a second nodal solution for a total of six nontrivial smooth solutions, all with precise sign information.

Our approach is variational based on the critical point theory, coupled with suitable truncation and comparison techniques and with Morse theory (critical groups). In the next section, for the convenience of the reader, we recall the main mathematical tools that we will use in this work.

2. Mathematical background

Let X be a Banach space. By X^* we denote the topological dual of X and by $\langle\cdot,\cdot\rangle$ the duality brackets for the pair (X^*, X). Let $\varphi\in C^1(X)$. We say that φ satisfies the *Cerami condition* if the following is true:

C-condition. Every sequence $\{x_n\}_{n\geq 1}\subseteq X$ such that $\{\varphi(x_n)\}_{n\geq 1}\subseteq\mathbb{R}$ is bounded and

$$(1+\|x_n\|)\varphi'(x_n)\to 0 \quad \text{in } X^* \text{ as } n\to\infty \tag{2}$$

admits a strongly convergent subsequence.

This compactness-type condition is in general weaker than the usual Palais–Smale condition ("PS-condition" for short). However, it suffices to have a deformation theorem and from it derive the minimax theory of certain critical values of φ (see, for example, [Gasiński and Papageorgiou 2006]). In particular, we can state the following theorem, known in the literature as the mountain pass theorem [ibid., p. 648].

Theorem 1. *If $\varphi\in C^1(X)$ satisfies the C-condition, $x_0, x_1\in X$, $\rho>0$, $\|x_0-x_1\|>\rho$,*

$$\max\{\varphi(x_0),\varphi(x_1)\}<\inf[\varphi(x):\|x-x_0\|=\rho]=\eta_\rho$$

and

$$c=\inf_{\gamma\in\Gamma}\max_{0\leq t\leq 1}\varphi(\gamma(t)),\quad \textit{where } \Gamma=\{\gamma\in C([0,1],X):\gamma(0)=x_0,\ \gamma(1)=x_1\},$$

then $c\geq\eta_\rho$ and c is a critical value of φ.

In the analysis of problem (1), in addition to the Sobolev spaces

$$W_0^{1,p}(\Omega), \quad H_0^1(\Omega),$$

we will also use the Banach space

$$C_0^1(\overline{\Omega}) = \{u \in C^1(\overline{\Omega}) : u|_{\partial\Omega} = 0\}.$$

This is an ordered Banach space with positive cone

$$C_+ = \{u \in C_0^1(\overline{\Omega}) : u(z) \geq 0 \text{ for all } z \in \Omega\}.$$

This cone has a nonempty interior given by

$$\operatorname{int} C_+ = \left\{u \in C_+ : u(z) > 0 \text{ for all } z \in \Omega,\ \frac{\partial u}{\partial n}(z) < 0 \text{ for all } z \in \partial\Omega\right\}$$

(here $n(\cdot)$ denotes the outward unit normal on $\partial\Omega$).

Let $f_0 : \Omega \times \mathbb{R} \to \mathbb{R}$ be a Carathéodory function with subcritical growth in $x \in \mathbb{R}$; i.e.,

$$|f_0(z, x)| \leq \hat{a}(z) + \hat{c}|x|^{r-1} \quad \text{for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R},$$

with $\hat{a} \in L^\infty(\Omega)_+$, $\hat{c} > 0$, and

$$1 < r < p^* = \begin{cases} \dfrac{Np}{N-p} & \text{if } p < N, \\ +\infty & \text{if } p \geq N. \end{cases}$$

We set $F_0(z, x) = \int_0^x f_0(z, s)\, ds$ and consider the C^1-functional $\varphi_0 : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_0(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega F_0(z, u(z))\, dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

The next theorem is a particular case of a more general result of [Gasiński and Papageorgiou 2012].

Theorem 2. *If $u_0 \in W_0^{1,p}(\Omega)$ is a local $C_0^1(\overline{\Omega})$-minimizer of φ_0, i.e., there exists $\rho_0 > 0$ such that*

$$\varphi_0(u_0) \leq \varphi_0(u_0 + h) \quad \textit{for all } h \in C_0^1(\overline{\Omega}) \textit{ with } \|h\|_{C_0^1(\overline{\Omega})} \leq \rho_0,$$

then $u_0 \in C^{1,\alpha}(\overline{\Omega})$ for some $\alpha \in (0, 1)$ and u_0 is also a local $W_0^{1,p}(\Omega)$-minimizer of φ_0; i.e., there exists $\rho_1 > 0$ such that

$$\varphi_0(u_0) \leq \varphi_0(u_0 + h) \quad \textit{for all } h \in W_0^{1,p}(\Omega) \textit{ with } \|h\| \leq \rho_1.$$

Remark. We should mention that the first such result was proved by Brézis and Nirenberg [1993] and was later extended by García Azorero, Peral Alonso, and Manfredi [García Azorero et al. 2000].

Let $h, \hat{h} \in L^\infty(\Omega)$. We write $h \prec \hat{h}$ if, for every compact $K \subseteq \Omega$, we can find $\varepsilon > 0$ such that

$$h(z) + \varepsilon \le \hat{h}(z) \quad \text{for a.a. } z \in K.$$

Clearly, if $h, \hat{h} \in C(\Omega)$ and $h(z) < \hat{h}(z)$ for all $z \in \Omega$, then $h \prec \hat{h}$. A straightforward modification of the proof of Proposition 2.6 of [Arcoya and Ruiz 2006] in order to accommodate the extra linear term $-\Delta u$ gives the following strong comparison principle.

Proposition 3. *If $\xi \ge 0$, $h, \hat{h} \in L^\infty(\Omega)$, $h \prec \hat{h}$, $u, v \in C_0^1(\overline{\Omega})$ are solutions of*

$$-\Delta_p u(z) - \Delta u(z) + \xi |u(z)|^{p-2} u(z) = h(z),$$
$$-\Delta_p v(z) - \Delta v(z) + \xi |v(z)|^{p-2} v(z) = \hat{h}(z) \quad \text{in } \Omega,$$

and $v \in \operatorname{int} C_+$, then $v - u \in \operatorname{int} C_+$.

Proof. We follow [Arcoya and Ruiz 2006] (see Proposition 2.6).

By nonlinear regularity, $u, v \in C^{1,\beta}(\overline{\Omega})$ $(0 < \beta < 1)$.

We have

$$A_p(u) + A(u) + \xi |u|^{p-2} u = h \le \hat{h} = A_p(v) + A(v) + \xi v^{p-1} \quad \text{in } W^{-1,p'}(\Omega).$$

Acting with $(u - v)^+ \in W_0^{1,p}(\Omega)$, we obtain

$$\big\langle A_p(u) - A_p(v), (u - v)^+ \big\rangle + \big\langle A(u) - A(v), (u - v)^+ \big\rangle + \int_\Omega \xi (|u|^{p-2} u - v^{p-1})(u - v)^+ \, dz \le 0,$$

which implies that $\|D(u - v)^+\|_2^2 \le 0$, since A_p is monotone; hence $u \le v$.

First we show that $u(z) \le v(z)$ for all $z \in \Omega$. For this purpose, we introduce

$$D_0 = \{z \in \Omega : u(z) = v(z)\} \quad \text{and} \quad D_1 = \{z \in \Omega : Du(z) = Dv(z) = 0\}.$$

We show that $D_0 \subseteq D_1$. So, let $z_0 \in D_0$. Since $u \le v$, the function $z \mapsto (u - v)(z)$ attains its maximum at $z_0 \in D_0$ and so we have $Du(z_0) = Dv(z_0)$. If $Du(z_0) \ne 0$, then we can find $\overline{B}_\rho(z_0) \subseteq \Omega$ such that

$$\|Du(z)\| > 0, \quad \|Dv(z)\| > 0, \quad (Du(z), Dv(z))_{\mathbb{R}^N} > 0 \quad \text{for all } z \in \overline{B}_\rho(z_0).$$

We set $w = v - u \in C_+ \setminus \{0\}$. Then w satisfies the linear elliptic equation

$$-\sum_{i,j=1}^N \frac{\partial}{\partial z_i}\left(\eta_{ij}(z) \frac{\partial w}{\partial z_j}\right) = -\xi (v^{p-1} - |u|^{p-2} u) + \hat{h} - h.$$

In this equation the coefficients $\eta_{ij}(\,\cdot\,)$ are given by

$$\eta_{ij}(z) = \delta_{ij}(\|Du(z)\|^{p-2}+1) + (p-2)\|Du(z)\|^{p-4}\frac{\partial u}{\partial z_i}(z)\frac{\partial u}{\partial z_j}(z)$$

for all $z \in \overline{B}_\rho(z_0)$ (see [Arcoya and Ruiz 2006, p. 854]). Hence $\eta_{ij} \in C^\beta(\overline{B}_\rho(z_0))$ with $\beta \in (0, 1)$ and the η_{ij} form a uniformly elliptic operator by taking $\rho \in (0, 1)$ even smaller if necessary. Then the strong maximum principle (see [Gilbarg and Trudinger 2001; Vázquez 1984]) implies that

$$u(z) < v(z) \quad \text{for all } z \in \overline{B}_\rho(z_0),$$

which contradicts the fact that $z_0 \in D_0$. So, we infer that $D_0 \subseteq D_1$.

Since by hypothesis $v \in \operatorname{int} C_+$, we see that D_1 is compact and so D_0 is compact. So, we can find $\Omega_1 \subseteq \Omega$ open and smooth such that

$$D_0 \subseteq \Omega_1 \subseteq \overline{\Omega}_1 \subseteq \Omega.$$

We can find $\varepsilon > 0$ such that

$$\begin{aligned} u(z) + \varepsilon &< v(z) \quad \text{for all } z \in \partial\Omega_1, \\ h(z) + \varepsilon &< \hat{h}(z) \quad \text{for a.a. } z \in \Omega_1. \end{aligned}$$

Let $\delta \in (0, \min\{\varepsilon, 1\})$ be such that

$$\xi\left||s|^{p-2}s - |s'|^{p-2}s'\right| < \varepsilon \quad \text{for all } s, s' \in [-\|u\|_\infty, \|v\|_\infty] \text{ with } |s-s'| < 2\delta.$$

Then we have

$$\begin{aligned} -\Delta_p(u+\delta) - \Delta(u+\delta) + \xi|u+\delta|^{p-2}(u+\delta) &= -\Delta_p(u) - \Delta(u) + \xi|u+\delta|^{p-2}(u+\delta) \\ &= \xi[|u+\delta|^{p-2}(u+\delta) - |u|^{p-2}u] + h \\ &\le h + \varepsilon \le \hat{h} = -\Delta_p v - \Delta v + \xi v^{p-1}, \end{aligned}$$

which implies $u + \delta \le v$ in Ω_1, by the weak maximum principle.

Since $D_0 \subseteq \Omega_1$, we infer that the boundary point theorem is valid for uniformly elliptic operators with Hölder continuous coefficients (see [Finn and Gilbarg 1957, Lemma 7, p. 31; Gilbarg and Trudinger 2001, p. 46]). So, for every $z_0 \in \partial\Omega$, we have

$$\frac{\partial w}{\partial n}(z_0) < 0,$$

and therefore $v - u \in \operatorname{int} C_+$. □

We now recall some basic facts concerning the spectrum of $(-\Delta_p, W_0^{1,p}(\Omega))$. We consider the nonlinear eigenvalue problem

$$\text{(3)} \qquad \begin{cases} -\Delta_p u(z) = \hat{\lambda}|u(z)|^{p-2}u(z) & \text{a.e. in } \Omega, \\ u|_{\partial\Omega} = 0. \end{cases}$$

A number $\hat{\lambda} \in \mathbb{R}$ is an eigenvalue of $(-\Delta_p, W_0^{1,p}(\Omega))$ if the problem (3) has a nontrivial solution $\hat{u} \in W_0^{1,p}(\Omega)$; that solution is an eigenfunction corresponding to the eigenvalue $\hat{\lambda}$. The smallest eigenvalue $\hat{\lambda}_1(p)$ of $(-\Delta_p, W_0^{1,p}(\Omega))$ has the following properties (see [Anane 1987; Anane and Tsouli 1996; García Azorero and Peral Alonso 1987]):

- $\hat{\lambda}_1(p)$ is positive and isolated.
- $\hat{\lambda}_1(p)$ is simple (its eigenspace is one-dimensional).
- $$\hat{\lambda}_1(p) = \inf\left[\frac{\|Du\|_p^p}{\|u\|_p^p} : u \in W_0^{1,p}(\Omega), u \not\equiv 0\right].$$

In this variational characterization of $\hat{\lambda}_1(p)$, the infimum is realized on the corresponding one-dimensional eigenspace. Moreover, it is clear from the third property above that the elements of the one-dimensional eigenspace do not change sign. In the sequel, by $\hat{u}_{1,p} \in W_0^{1,p}(\Omega)$, we denote the L^p-normalized (i.e., $\|\hat{u}_{1,p}\|_p = 1$) positive eigenfunction corresponding to the eigenvalue $\hat{\lambda}_1(p) > 0$. The nonlinear regularity theory (see, for example, [Gasiński and Papageorgiou 2006, pp. 737–738]), implies that $\hat{u}_{1,p} \in C_+ \setminus \{0\}$. Then the nonlinear maximum principle of [Vázquez 1984] says that $\hat{u}_{1,p} \in \operatorname{int} C_+$. Since the spectrum $\sigma(p)$ of $(-\Delta_p, W_0^{1,p}(\Omega))$ is closed and $\hat{\lambda}_1(p) > 0$ is isolated, the second eigenvalue $\hat{\lambda}_2(p) = \inf[\lambda \in \sigma(p) : \lambda > \hat{\lambda}_1(p)]$ is also well-defined.

If $N = 1$ (ordinary differential equation), then $\sigma(p) = \{\hat{\lambda}_k(p)\}_{k\geq 1} \subseteq (0, +\infty)$, where each $\hat{\lambda}_k(p)$ is a simple eigenvalue, $\hat{\lambda}_k(p) \to +\infty$ as $k \to +\infty$ and the corresponding eigenfunctions $\{\hat{u}_{k,p}\}_{k\geq 1}$ have exactly $k-1$ zeros (see, for example, [Gasiński and Papageorgiou 2006, p. 761]).

If $N \geq 2$ (partial differential equation), then the Ljusternik–Schnirelmann minimax scheme via the Krasnoselskii genus gives us a whole strictly increasing sequence of eigenvalues $\{\hat{\lambda}_k(p)\}_{k\geq 1}$ such that $\hat{\lambda}_k(p) \to +\infty$ as $k \to +\infty$. It is not known if this is the complete list of eigenvalues. If $p = 2$ (linear eigenvalue problem), then these are all the eigenvalues of $(-\Delta, H_0^1(\Omega))$.

Next we recall some basic definitions and facts from Morse theory and from [Cingolani and Vannella 2003; 2007], which we will need in order to produce a second nodal solution.

So, as before, let X be a Banach space and (Y_1, Y_2) a topological pair such that $Y_2 \subseteq Y_1 \subseteq X$. For every integer $k \geq 0$, by $H_k(Y_1, Y_2)$ we denote the k-th-relative singular homology group with integer coefficients for the pair (Y_1, Y_2). For $k < 0$, $H_k(Y_1, Y_2) = 0$.

Given $\varphi \in C^1(X)$ and $c \in \mathbb{R}$, we introduce the sets

$$\varphi^c = \{x \in X : \varphi(x) \leq c\}, \quad K_\varphi = \{x \in X : \varphi'(x) = 0\}, \quad K_\varphi^c = \{x \in K_\varphi : \varphi(x) = c\}.$$

The critical groups of $\varphi \in C^1(X)$ at an isolated critical point $x \in X$ with $\varphi(x) = c$ (i.e., $x \in K_\varphi^c$) are defined by

$$C_k(\varphi, x) = H_k\big(\varphi^c \cap U, \varphi^c \cap U \setminus \{x\}\big) \quad \text{for all } k \geq 0,$$

where U is a neighborhood of x such that $K_\varphi \cap \varphi^c \cap U = \{x\}$. The excision property of singular homology theory implies that the above definition of critical groups is independent of the particular choice of the neighborhood U.

Now suppose that $\varphi \in C^1(X)$ satisfies the C-condition and $\inf \varphi(K_\varphi) > -\infty$. Let $c < \inf \varphi(K_\varphi)$. The critical groups of φ at infinity are defined by

$$C_k(\varphi, \infty) = H_k(X, \varphi^c) \quad \text{for all } k \geq 0.$$

The second deformation theorem (see, for example, [Gasiński and Papageorgiou 2006, p. 628]), implies that this definition is independent of the level $c < \inf \varphi(K_\varphi)$.

Suppose that K_φ is finite and define

$$M(t, x) = \sum_{k \geq 0} \operatorname{rank} C_k(\varphi, x) t^k \quad \text{for all } t \in \mathbb{R}, \text{ all } x \in K_\varphi$$

and

$$P(t, \infty) = \sum_{k \geq 0} \operatorname{rank} C_k(\varphi, \infty) t^k \quad \text{for all } t \in \mathbb{R}.$$

The Morse relation says that

$$\sum_{x \in K_\varphi} M(t, x) = P(t, \infty) + (1 + t) Q(t), \tag{4}$$

where $Q(t) = \sum_{k \geq 0} \beta_k t^k$ is a formal series in $t \in \mathbb{R}$ with integer coefficients β_k.

Let H be a Hilbert space, x a point in H, and U a neighborhood of x. Let $\varphi \in C^2(U)$. If $x \in K_\varphi$, then the Morse index $\mu = \mu(x)$ of x is defined to be the supremum of the dimensions of the vector subspaces of H on which $\varphi''(x)$ is negative definite. The nullity $\nu(x)$ of $x \in K_\varphi$ is the dimension of $\ker \varphi''(x)$. We say that $x \in K_\varphi$ is nondegenerate if $\varphi''(x)$ is invertible (i.e., $\nu(x) = 0$). If $\varphi \in C^2(U)$ and $x \in K_\varphi$ is nondegenerate with Morse index μ, then

$$C_k(\varphi, x) = \delta_{k,\mu} \mathbb{Z} \quad \text{for all } k \geq 0,$$

where $\delta_{k,\mu}$ is the Kronecker symbol.

As mentioned in the introduction, to produce a second nodal solution, we will use some facts from [Cingolani and Vannella 2003; 2007]. Suppose $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a measurable function such that, for a.a. $z \in \Omega$, $f(z, \cdot) \in C^1(\mathbb{R})$ and

$$|f_x'(z, x)| \leq \tilde{\alpha}(z) + \tilde{c}|x|^{r-2} \quad \text{for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R},$$

with $\tilde{\alpha} \in L^\infty(\Omega)_+$, $\tilde{c} > 0$ and $p \le r < p^*$. We set $F(z,x) = \int_0^x f(z,s)\,ds$ and consider the C^2-functional $\varphi : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega F(z,u(z))\,dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

For all $u, v, y \in W_0^{1,p}(\Omega)$, we have (see [Cingolani and Vannella 2003])

$$\langle \varphi''(u)v, y\rangle = \int_\Omega (1+\|Du\|^{p-2})(Dv, Dy)_{\mathbb{R}^N}\,dz \\ + (p-2)\int_\Omega \|Du\|^{p-4}(Du, Dv)_{\mathbb{R}^N}(Du, Dy)_{\mathbb{R}^N}\,dz - \int_\Omega f'_x(z,u)vy\,dz.$$

Here $\langle\cdot,\cdot\rangle$ denotes the duality brackets for the pair consisting of the spaces

$$W^{-1,p'}(\Omega) = W_0^{1,p}(\Omega)^* \quad \text{and} \quad W_0^{1,p}(\Omega), \quad \text{where } \frac{1}{p} + \frac{1}{p'} = 1.$$

Suppose that $u_0 \in K_\varphi$. Nonlinear regularity theory (see [Ladyzhenskaya and Ural'tseva 1968; Lieberman 1991]) implies that $u_0 \in C_0^1(\overline{\Omega})$. It follows that

$$b(\cdot) = \|Du_0(\cdot)\|^{(p-4)/2} Du_0(\cdot) \in L^\infty(\Omega, \mathbb{R}^N).$$

Let H_b be the completion of $C_c^\infty(\Omega)$ under the inner product

$$(v,y)_b = \int_\Omega \big[(1+\|b\|^2)(Dv,Dy)_{\mathbb{R}^N} + (p-2)(b,Dv)_{\mathbb{R}^N}(b,Dy)_{\mathbb{R}^N}\big]\,dz.$$

Denote by $\|\cdot\|_b$ the corresponding norm. Clearly $\|\cdot\|_b$ is equivalent to the usual Sobolev norm of $H_0^1(\Omega)$, so H_b and $H_0^1(\Omega)$ are isomorphic. Since $p > 2$, $W_0^{1,p}(\Omega)$ is embedded continuously into H_b. Let $L_b \in \mathcal{L}(H_b, H_b^*)$ be defined by

$$\langle L_b(v), y\rangle_b = (v,y)_b - \int_\Omega f'_x(z,u_0)vy\,dz \quad \text{for all } v, y \in H_b.$$

Then L_b is a Fredholm operator of index zero and it is the extension of $\varphi''(u_0)$ on H_b. We consider the orthogonal direct sum decomposition

$$H_b = H^- \oplus H^0 \oplus H^+,$$

where H^-, H^0, H^+ are the negative, null and positive spaces according to the spectral decomposition of L_b in $L^2(\Omega)$. Then H^- and H^0 are finite-dimensional and, since $u_0 \in C_0^1(\overline{\Omega})$, standard regularity theory implies that

$$H^- \oplus H^0 \subseteq W_0^{1,p}(\Omega) \cap L^\infty(\Omega).$$

We set $V = H^- \oplus H^0$ and $W = W_0^{1,p}(\Omega) \cap H^+$. Then $W_0^{1,p}(\Omega) = V \oplus W$ and, by [Cingolani and Vannella 2003, p. 279], there exists $c > 0$ such that

$$\langle \varphi''(u_0)v, v\rangle \ge c\|v\|_b^2 \quad \text{for all } v \in W.$$

In what follows, for every $r \in (1, +\infty)$, we denote by

$$A_r : W_0^{1,r}(\Omega) \to W^{-1,r'}(\Omega), \qquad \frac{1}{r} + \frac{1}{r'} = 1,$$

the nonlinear map defined by

$$\langle A_r(u), y\rangle = \int_\Omega \|Du\|^{r-2}(Du, Dy)_{\mathbb{R}^N}\, dz \quad \text{for all } u, y \in W_0^{1,r}(\Omega). \tag{5}$$

If $r = 2$, then we set $A_2 = A \in \mathscr{L}(H_0^1(\Omega), H^{-1}(\Omega))$. The next result can be found in [Gasiński and Papageorgiou 2006, pp. 745–746].

Proposition 4. *If $A_r : W_0^{1,r}(\Omega) \to W^{-1,r'}(\Omega)$ is defined by* (5), *then A_r is continuous, monotone* (*hence maximal monotone*) *and of type* $(S)_+$; *that is, if u_n converges weakly to u in $W_0^{1,p}(\Omega)$ and* $\limsup_{n\to+\infty}\langle A_r(u_n), u_n - u\rangle \le 0$, *then $u_n \to u$ in $W_0^{1,r}(\Omega)$.*

Throughout this paper by $\|\cdot\|$ we denote the norm of $W_0^{1,p}(\Omega)$. By virtue of Poincaré's inequality, $\|u\| = \|Du\|_p$ for all $u \in W_0^{1,p}(\Omega)$. By $\|\cdot\|$ we will also denote the norm of $\mathbb{R}^N$. No confusion is possible, since it will always be clear from the context which norm we mean.

For $x \in \mathbb{R}$, we define

$$x^\pm = \max\{\pm x, 0\}.$$

Then for $u \in W_0^{1,p}(\Omega)$ we set

$$u^\pm(\,\cdot\,) = u(\,\cdot\,)^\pm.$$

We know that $u^\pm \in W_0^{1,p}(\Omega)$ and

$$|u| = u^+ + u^- \quad \text{and} \quad u = u^+ - u^-.$$

By $|\cdot|_{\mathbb{R}^N}$ we denote the Lebesgue measure on $\mathbb{R}^N$.

Finally, if $g : \Omega\times\mathbb{R} \to \mathbb{R}$ is a measurable function (for example, if $(z, x) \to g(z, x)$ is a Carathéodory function), then we set

$$N_g(u)(\,\cdot\,) = g(\,\cdot\,, u(\,\cdot\,)) \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

3. Constant sign solutions

In this section, we produce four nontrivial smooth solutions of constant sign, two positive and two negative. The hypotheses on the reaction $f(z, x)$ are the following:

Hypotheses H. (i) $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a Carathéodory function.

(ii) $f(z, 0) = 0$ a.e. in Ω.

(iii) $|f(z, x)| \le \alpha(z) + c|x|^{p-1}$ for a.a. $z \in \Omega$, all $x \in \mathbb{R}$, with $\alpha \in L^\infty(\Omega)_+$, $c > 0$.

(iv) If $F(z,x)=\int_0^x f(z,s)\,ds$, then

$$\lim_{x\to\pm\infty}\frac{pF(z,x)}{|x|^p}=\hat{\lambda}_1(p)\quad\text{uniformly for a.a. } z\in\Omega,$$

and, for some $\tau>2$,

$$\limsup_{x\to\pm\infty}\frac{f(z,x)x-pF(z,x)}{|x|^\tau}\le\hat{\beta}<0\quad\text{uniformly for a.a. } z\in\Omega.$$

(v) There exist functions $w_\pm\in W^{1,p}(\Omega)\cap C(\overline{\Omega})$ such that

$$w_-(z)\le c_-<0<c_+\le w_+(z)\quad\text{for all } z\in\overline{\Omega},$$
$$\operatorname*{ess\,sup}_{\Omega} f(\,\cdot\,,w_+(\,\cdot\,))\le 0\le\operatorname*{ess\,inf}_{\Omega} f(\,\cdot\,,w_-(\,\cdot\,)),$$

and $A_p(w_-)+A(w_-)\le 0\le A_p(w_+)+A(w_+)$ in $W^{-1,p'}(\Omega)=W_0^{1,p}(\Omega)^*$.

(vi) For every $\rho>0$, there exists $\xi_\rho>0$ such that for a.a. $z\in\Omega$, the function $x\to f(z,x)+\xi_\rho|x|^{p-2}x$ is nondecreasing on $[-\rho,\rho]$.

(vii) There exist integer $m\ge 2$ and functions $\eta,\hat{\eta}\in L^\infty(\Omega)_+$ such that

$$\hat{\lambda}_m(2)\le\eta(z)\le\hat{\eta}(z)\le\hat{\lambda}_{m+1}(2)\quad\text{a.e. in }\Omega,\quad\hat{\lambda}_m(2)\ne\eta,\quad\hat{\lambda}_{m+1}(2)\ne\hat{\eta}$$

and

$$\eta(z)\le\liminf_{x\to 0}\frac{f(z,x)}{x}\le\limsup_{x\to 0}\frac{f(z,x)}{x}\le\hat{\eta}(z)\quad\text{uniformly for a.a. } z\in\Omega.$$

Remarks. Hypothesis H(iv) implies that, asymptotically at $\pm\infty$, we have resonance with respect to the principal eigenvalue $\hat{\lambda}_1(p)>0$ from the right. Hence the energy functional of the problem, as we will see, is indefinite. Hypothesis H(v) is satisfied if we can find $c_-<0<c_+$ such that $f(z,c_+)=f(z,c_-)=0$ a.e. in Ω.

Example. The following function satisfies the hypotheses H (for simplicity, we drop the z-dependence):

$$f(x)=\begin{cases}\eta(x-|x|^{r-2}x) & \text{if } |x|\le 1,\\ \hat{\lambda}_1(p)(|x|^{p-2}x-|x|^{\tau-2}x) & \text{if } |x|>1,\end{cases}$$

with $\eta\in(\hat{\lambda}_m(2),\hat{\lambda}_{m+1}(2))$, $m\ge 2$ and $r>2$, $1<\tau<p$.

We introduce the following truncations of $f(z,\cdot)$:

$$\hat{f}_+(z,x)=\begin{cases}0 & \text{if } x<0,\\ f(z,x) & \text{if } 0\le x\le w_+(z),\\ f(z,w_+(z)) & \text{if } w_+(z)<x\end{cases}\tag{6}$$

and

$$\hat{f}_-(z,x) = \begin{cases} f(z, w_-(z)) & \text{if } x < w_-(z), \\ f(z,x) & \text{if } w_-(z) \le x \le 0, \\ 0 & \text{if } 0 < x. \end{cases}$$

Both are Carathéodory functions. Let $\hat{F}_\pm(z,x) = \int_0^x \hat{f}_\pm(z,s)\,ds$ and consider the C^1-functionals $\hat{\varphi}_\pm : W_0^{1,p}(Z) \to \mathbb{R}$ defined by

$$\hat{\varphi}_\pm(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega \hat{F}_\pm(z,u(z))\,dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

Also, let $\varphi : W_0^{1,p}(\Omega) \to \mathbb{R}$ be the energy (Euler) functional for problem (1) defined by

$$\varphi(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega F(z,u(z))\,dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

Clearly, $\varphi \in C^1(W_0^{1,p}(\Omega))$.

First, we produce two nontrivial constant sign smooth solutions of (1).

Proposition 5. *If hypotheses* H(iii), (v), (vi), (vii) *hold*, *then problem* (1) *has at least the two nontrivial constant sign smooth solutions*

$$u_0 \in \operatorname{int} C_+ \quad \textit{and} \quad v_0 \in -\operatorname{int} C_+,$$

and both are local minimizers of φ.

Proof. First we produce the positive solution.

From (6) we see that $\hat{\varphi}_+$ is coercive. Also, using the Sobolev embedding theorem, we can check easily that $\hat{\varphi}_+$ is sequentially weakly lower semicontinuous. So, by the Weierstrass theorem, we can find $u_0 \in W_0^{1,p}(\Omega)$ such that

$$\hat{\varphi}_+(u_0) = \inf[\hat{\varphi}_+(u) : u \in W_0^{1,p}(\Omega)] = m_+. \tag{7}$$

By virtue of hypothesis H(vii), we can find $\vartheta > \hat{\lambda}_1(2)$ and $0 < \delta < \min\{c_+, -c_-\}$ such that

$$F(z,x) \ge \tfrac{1}{2}\vartheta x^2 \quad \text{for a.a. } z \in \Omega, \text{ all } |x| \le \delta. \tag{8}$$

Let $t \in (0,1)$ be small such that $t\hat{u}_{1,2}(z) \in [0,\delta]$ for all $z \in \overline{\Omega}$ (recall that $\hat{u}_{1,2} \in \operatorname{int} C_+$). Then

$$\begin{aligned} \hat{\varphi}_+(t\hat{u}_{1,2}) &= \frac{t^p}{p}\|D\hat{u}_{1,2}\|_p^p + \frac{t^2}{2}\hat{\lambda}_1(2) - \int_\Omega \hat{F}_+(z, t\hat{u}_{1,2})\,dz \\ &\le \frac{t^p}{p}\|D\hat{u}_{1,2}\|_p^p + \frac{t^2}{2}[\hat{\lambda}_1(2) - \vartheta] \end{aligned}$$

(see (8) and recall that $\|\hat{u}_{1,2}\|_2 = 1$).

Since $\vartheta > \hat{\lambda}_1(2)$ and $p > 2$, by choosing $t \in (0, 1)$ even smaller if necessary, we have $\hat{\varphi}_+(t\hat{u}_{1,2}) < 0$, which implies $\hat{\varphi}_+(u_0) = \hat{m}_+ < 0 = \hat{\varphi}_+(0)$ (see (7)); hence $u_0 \neq 0$.

From (7) we have $\hat{\varphi}'_+(u_0) = 0$, which implies

$$A_p(u_0) + A(u_0) = N_{\widehat{f}_+}(u_0). \tag{9}$$

On (9) we act with $-u_0^- \in W_0^{1,p}(\Omega)$ and obtain $u_0 \geq 0$, $u_0 \neq 0$ (see (6)). Also, we act with $(u_0 - w_+)^+ \in W_0^{1,p}(\Omega)$. Then

$$\begin{aligned}\langle A_p(u_0), (u_0 - w_+)^+\rangle + \langle A(u_0), (u_0 - w_+)^+\rangle \\ &= \int_\Omega \hat{f}_+(z, u_0)(u_0 - w_+)^+\,dz = \int_\Omega f(z, w_+)(u_0 - w_+)^+\,dz \quad \text{(see (6))} \\ &\leq \big\langle A_p(w_+) + A(w_+), (u_0 - w_+)^+\big\rangle,\end{aligned}$$

by hypothesis H(v). Therefore

$$\int_{\{u_0 > w_+\}} \big(\|Du_0\|^{p-2}Du_0 - \|Dw_+\|^{p-2}Dw_+, Du_0 - Dw_+\big)_{\mathbb{R}^N}\,dz \\ + \|D(u_0 - w_+)^+\|_2^2 \leq 0.$$

It follows that $u_0 \leq w_+$.

So, we have proved that

$$u_0 \in [0, w_+] = \{u \in W_0^{1,p}(\Omega) : 0 \leq u(z) \leq w_+(z) \text{ a.e. in } \Omega\}.$$

Then (9) becomes $A_p(u_0) + A(u_0) = N_f(u_0)$ (see (6)), and hence

$$-\Delta_p u_0(z) - \Delta u_0(z) = f(z, u_0(z)) \quad \text{a.e. in } \Omega, \quad u_0|_{\partial\Omega} = 0. \tag{10}$$

From (10) and [Ladyzhenskaya and Ural'tseva 1968, Theorem 7.1, p. 286], we have $u_0 \in L^\infty(\Omega)$. We can apply the regularity result of [Lieberman 1991, p. 320] and have $u_0 \in C_+ \setminus \{0\}$. Note that

$$A_p(u_0) + A(u_0) - N_f(u_0) = 0 \leq A_p(w_+) + A(w_+) - N_f(w_+) \quad \text{in } W^{-1,p'}(\Omega),$$

by H(v), and, for a.a. $z \in \Omega$ and all $x, y \in [-\rho, \rho]$ with $x > y$, we have, by H(vi),

$$f(z, x) - f(z, y) \geq -\xi_\rho(x - y),$$

Let $a(\xi) = \|\xi\|^{p-2}\xi + \xi$ for all $\xi \in \mathbb{R}^N$. Then $a \in C^1(\mathbb{R}^N)$,

$$\nabla a(\xi) = \|\xi\|^{p-2}\bigg(I + (p-2)\frac{\xi \otimes \xi}{\|\xi\|^2}\bigg) + I$$

and

$$\operatorname{div} a(Du) = \Delta_p u + \Delta u \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

We have $(\nabla a(\xi)y, y)_{\mathbb{R}^N} \geq \|y\|^2$ for all ξ, $y \in \mathbb{R}^N$ and so we can apply [Pucci and Serrin 2007, Theorem 2.5.3, p. 37] and infer, via H(v), that

$$u_0(z) < w_+(z) \quad \text{for all } z \in \overline{\Omega}.$$

Let $\rho = \max\{\|w_+\|_\infty, \|w_-\|_\infty\}$. By virtue of H(vi) and (10), we have

$$-\Delta_p u_0(z) - \Delta u_0(z) + \hat{\xi}_\rho u_0(z)^{p-1} = f(z, u_0(z)) + \hat{\xi}_\rho u_0(z)^{p-1} \geq 0 \quad \text{a.e. in } \Omega,$$

and hence

$$\Delta_p u_0(z) + \Delta u_0(z) \leq \hat{\xi}_\rho u_0(z)^{p-1} \quad \text{a.e. in } \Omega.$$

Invoking the boundary point theorem of Pucci and Serrin [2007, Theorem 5.5.1, p. 120] we have $u_0 \in \operatorname{int} C_+$. Therefore

$$u_0 \in \operatorname{int}_{C_0^1(\overline{\Omega})}[0, w_+].$$

It is clear from (6) that $\hat{\varphi}_+|_{[0,w_+]} = \varphi|_{[0,w_+]}$. Therefore u_0 is a local $C_0^1(\overline{\Omega})$-minimizer of φ and so by Theorem 2 it is also a local $W_0^{1,p}(\Omega)$-minimizer of φ.

Similarly, working this time with $\hat{\varphi}_-$, we produce another constant sign smooth solution $v_0 \in -\operatorname{int} C_+$ which is a local minimizer of φ. $\square$

Using $u_0 \in \operatorname{int} C_+$, $v_0 \in -\operatorname{int} C_+$, we can produce two more nontrivial constant sign smooth solutions.

Proposition 6. *If hypotheses* H *hold and* K_φ *is finite*, *problem* (1) *has at least four nontrivial constant sign smooth solutions*

$$u_0, \hat{u} \in \operatorname{int} C_+ \quad \textit{with } \hat{u} - u_0 \in \operatorname{int} C_+$$
$$v_0, \hat{v} \in -\operatorname{int} C_+ \quad \textit{with } v_0 - \hat{v} \in \operatorname{int} C_+.$$

Proof. From Proposition 5 we already have two solutions $u_0 \in \operatorname{int} C_+$ and $v_0 \in -\operatorname{int} C_+$.

Next we produce the second nontrivial positive smooth solution. To this end, we introduce the following truncation of $f(z, \cdot)$:

$$h_+(z, x) = \begin{cases} f(z, u_0(z)) & \text{if } x \leq u_0(z), \\ f(z, x) & \text{if } u_0(z) < x. \end{cases} \tag{11}$$

This is a Carathéodory function. We set $H_+(z, x) = \int_0^x h_+(z, s)\, ds$ and consider the C^1-functional $\psi_+ : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\psi_+(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega H_+(z, u(z))\, dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

Claim 1. *The functional* ψ_+ *satisfies the C-condition.*

Proof. Let $\{u_n\}_{n\geq 1} \subseteq W_0^{1,p}(\Omega)$ be a sequence such that

$$|\psi_+(u_n)| \leq M_1 \quad \text{for some } M_1 > 0, \text{ all } n \geq 1 \tag{12}$$

and

$$(1+\|u_n\|)\psi_+'(u_n) \to 0 \quad \text{in } W^{-1,p'}(\Omega) \text{ as } n \to \infty. \tag{13}$$

From (13) we have

$$\left| \langle A_p(u_n), g\rangle + \langle A(u_n), g\rangle - \int_\Omega h_+(z, u_n) g \, dz \right| \leq \frac{\varepsilon_n \|g\|}{1+\|u_n\|} \tag{14}$$

for all $g \in W_0^{1,p}(\Omega)$, with $\varepsilon_n \downarrow 0$.

In (14) we choose $g = -u_n^- \in W_0^{1,p}(\Omega)$. Then we get

$$\|Du_n^-\|_p^p + \|Du_n^-\|_2^2 - \int_\Omega f(z, u_0)(-u_n^-)\, dz \leq \varepsilon_n \quad \text{for all } n \geq 1,$$

by (11); this implies that $\|Du_n^-\|_p^p \leq c_1 \|u_n^-\|$ for some $c_1 > 0$ and all $n \geq 1$ (by H(iii)), and we conclude, since $p > 1$, that

$$\{u_n^-\}_{n\geq 1} \subseteq W_0^{1,p}(\Omega) \quad \text{is bounded.} \tag{15}$$

We will show that $\{u_n\}_{n\geq 1} \subseteq W_0^{1,p}(\Omega)$ is bounded. Arguing by contradiction, because of (15) and by passing to a suitable subsequence if necessary, we may assume that $\|u_n^+\| \to \infty$. We set $y_n = u_n^+/\|u_n^+\|$, $n \geq 1$. Then $\|y_n\| = 1$ for all $n \geq 1$ and so we may assume that

$$y_n \xrightarrow{w} y \text{ in } W_0^{1,p}(\Omega) \quad \text{and} \quad y_n \to y \text{ in } L^p(\Omega) \quad \text{as } n \to \infty, \tag{16}$$

where $\xrightarrow{w}$ indicates weak convergence. From (14), we have

$$\left| \langle A_p(y_n), g\rangle + \frac{1}{\|u_n^+\|^{p-2}} \langle A(y_n), g\rangle - \int_\Omega \frac{h_+(z, u_n^+)}{\|u_n^+\|^{p-1}} g \, dz \right| \leq \varepsilon_n' \|g\|, \tag{17}$$

with $\varepsilon_n' \to 0$ (see (15)).

Hypothesis H(iii) and (11) imply that

$$\left\{ \frac{N_{h_+}(u_n^+)}{\|u_n^+\|^{p-1}} \right\}_{n\geq 1} \subseteq L^{p'}(\Omega) \quad \text{is bounded.} \tag{18}$$

From (18) and using hypothesis H(iv), as in the proof of Proposition 30 of [Aizicovici et al. 2008], we have

$$\frac{N_{h_+}(u_n^+)}{\|u_n^+\|^{p-1}} \xrightarrow{w} \beta = \hat{\lambda}_1(p) y^{p-1} \quad \text{in } L^{p'}(\Omega). \tag{19}$$

Also, if in (17) we choose $g = y_n - y \in W_0^{1,p}(\Omega)$, pass to the limit as $n \to \infty$ and use (16) and (19), we obtain

$$\lim_{n\to\infty} [\langle A_p(y_n), y_n - y\rangle + \langle A(y_n), y_n - y\rangle] = 0,$$

from which we get successively

$$\limsup_{n\to\infty} [\langle A_p(y_n), y_n - y\rangle + \langle A(y), y_n - y\rangle] \le 0 \quad \text{(since } A \text{ is monotone)},$$

$$\limsup_{n\to\infty} \langle A_p(y_n), y_n - y\rangle \le 0 \quad \text{(see (16))},$$

$$y_n \to y \quad \text{in } W_0^{1,p}(\Omega) \qquad \text{(see Proposition 4)}.$$

The upshot is that

$$\|y\| = 1, \quad y \ge 0. \tag{20}$$

Passing to the limit as $n \to \infty$ in (17) and using (19) and (20), we see that

$$\langle A_p(y), g\rangle = \hat{\lambda}_1(p) \int_\Omega y^{p-1} g\, dz \quad \text{for all } g \in W_0^{1,p}(\Omega),$$

since $p > 2$ and $\|u_n^+\| \to \infty$. This yields $A_p(y)\hat{\lambda}_1(p)y^{p-1}$ and so

$$-\Delta_p y(z) = \hat{\lambda}_1(p) y(z)^{p-1} \quad \text{a.e. in } \Omega, \quad y|_{\partial\Omega} = 0,$$

implying, in view of (20), that

$$y = \lambda \hat{u}_{1,p} \quad \text{for some } \lambda > 0. \tag{21}$$

Therefore $y(z) > 0$ for all $z \in \Omega$ and this implies that $u_n^+(z) \to +\infty$ for all $z \in \Omega$. Then, by virtue of hypothesis H(iv), we have

$$\limsup_{n\to\infty} \frac{f(z, u_n^+(z))u_n^+(z) - pF(z, u_n^+(z))}{|u_n^+(z)|^\tau} \le \hat{\beta} < 0 \quad \text{for a.a. } z \in \Omega,$$

or again, in view of (11),

$$\limsup_{n\to\infty} \frac{h_+(z, u_n^+(z))u_n^+(z) - pH_+(z, u_n^+(z))}{|u_n^+(z)|^\tau} \le \hat{\beta} < 0 \quad \text{for a.a. } z \in \Omega. \tag{22}$$

Hypothesis H(iv) and Fatou's lemma, together with (21) and (22), imply that

$$\limsup_{n\to\infty} \frac{1}{\|u_n^+\|^\tau} \int_\Omega \big[h_+(z, u_n^+)u_n^+(z) - pH_+(z, u_n^+)\big]\, dz < 0 \tag{23}$$

On the other hand, from (12) and (15), we have

$$-M_2 \le \|Du_n^+\|_p^p + \frac{p}{2}\|Du_n^+\|_2^2 - \int_\Omega pH_+(z, u_n^+)\, dz \tag{24}$$

for some $M_2 > 0$ and all $n \geq 1$.

Also, if we choose $g = u_n^+ \in W_0^{1,p}(\Omega)$ in (14), then

$$-\varepsilon_n \leq -\|Du_n^+\|_p^p - \|Du_n^+\|_2^2 + \int_\Omega h_+(z, u_n^+)u_n^+ \, dz \quad \text{for all } n \geq 1. \tag{25}$$

Adding (24) and (25), we obtain

$$-M_3 \leq \int_\Omega \big[h_+(z, u_n^+)u_n^+ - pH_+(z, u_n^+)\big] \, dz + \Big(\frac{p}{2} - 1\Big)\|Du_n^+\|_2^2$$

for some $M_3 > 0$ and all $n \geq 1$, whence (since $p > 2$)

$$-\frac{M_3}{\|u_n^+\|^\tau} \leq \frac{1}{\|u_n^+\|^\tau} \int_\Omega \big[h_+(z, u_n^+)u_n^+ - pH_+(z, u_n^+)\big] \, dz + c_2\Big(\frac{p}{2} - 1\Big)\frac{1}{\|u_n^+\|^{\tau-2}}$$

for some $c_2 > 0$ and all $n \geq 1$, and finally, since $\tau > 2$ and $p > 2$,

$$0 \leq \liminf_{n\to\infty} \frac{1}{\|u_n^+\|^\tau} \int_\Omega \big[h_+(z, u_n^+)u_n^+ - pH_+(z, u_n^+)\big] \, dz. \tag{26}$$

Comparing (23) and (26), we reach a contradiction.

This proves that $\{u_n^+\}_{n\geq 1} \subseteq W_0^{1,p}(\Omega)$ is bounded; hence $\{u_n\}_{n\geq 1} \subseteq W_0^{1,p}(\Omega)$ is bounded, by (15). So, we may assume that

$$u_n \xrightarrow{w} u \text{ in } W_0^{1,p}(\Omega) \quad \text{and} \quad u_n \to u \text{ in } L^p(\Omega). \tag{27}$$

If in (14) we choose $g = u_n - u \in W_0^{1,p}(\Omega)$, pass to the limit as $n \to \infty$ and use (27), then, as before, exploiting the monotonicity of A, we have

$$\limsup_{n\to\infty} \langle A_p(u_n), u_n - u\rangle \leq 0,$$

implying that $u_n \to u$ in $W_0^{1,p}(\Omega)$, by Proposition 4. Hence ψ_+ satisfies the C-condition, and this proves Claim 1. □

Claim 2. *The function u_0 is a local minimizer of ψ_+.*

Proof. We may assume that $K_\varphi \cap [0, w_+] = \{0, u_0\}$. Otherwise, let y be a nontrivial element of $K_\varphi \cap [0, w_+]$ distinct from u_0; as a nontrivial solution of (1), y can be taken such that $u_0 \leq y$, because (1) has a biggest solution in $[0, w_+]$ (this is shown like Proposition 4.4 in [Filippakis et al. 2009]). Therefore, we are done if such a y exists.

We introduce the following truncation of $h_+(z, \cdot)$:

$$\hat{h}_+(z, x) = \begin{cases} h_+(z, x) & \text{if } x \leq w_+(z), \\ h_+(z, w_+(z)) & \text{if } w_+(z) < x. \end{cases} \tag{28}$$

This is a Carathéodory function. We set $\hat{H}_+(z, x) = \int_0^x \hat{h}_+(z, s) \, ds$ and consider the C^1-functional $\hat{\psi}_+ : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\hat{\psi}_+(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega \hat{H}_+(z, u(z))\,dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

From (28) it is clear that $\hat{\psi}_+$ is coercive. Also, it is sequentially weakly lower semicontinuous. Hence, we can find $\bar{u}_0 \in W_0^{1,p}(\Omega)$ such that

$$\hat{\psi}_+(\bar{u}_0) = \inf\,[\hat{\psi}_+(u) : u \in W_0^{1,p}(\Omega)],$$

which is to say $\hat{\psi}'_+(\bar{u}_0) = 0$; therefore

$$A_p(\bar{u}_0) + A(\bar{u}_0) = N_{\widehat{h}_+}(\bar{u}_0). \tag{29}$$

On (29) first we act with $(u_0 - \bar{u}_0)^+ \in W_0^{1,p}(\Omega)$ and then with $(\bar{u}_0 - w_+)^+ \in W_0^{1,p}(\Omega)$. Using (11), (28) and hypothesis H(v), this leads to

$$\bar{u}_0 \in [u_0, w_+] = \big\{u \in W_0^{1,p}(\Omega) : u_0(z) \le u(z) \le w_+(z) \text{ a.e. in } \Omega\big\}.$$

Then (29) becomes $A_p(\bar{u}_0) + A(\bar{u}_0) = N_f(\bar{u}_0)$ by (11) and (28); thus $\bar{u}_0 \in K_\varphi \cap [0, w_+]$, which is to say $\bar{u}_0 = u_0$.

From Proposition 5 and its proof, we have

$$u_0 \in \operatorname{int} C_+ \quad \text{and} \quad u_0(z) < w_+(z) \quad \text{for all } z \in \overline{\Omega}.$$

From (28) we infer that $\psi_+|_{[0,w_+]} = \hat{\psi}_+|_{[0,w_+]}$, so u_0 is a local $C_0^1(\overline{\Omega})$-minimizer of ψ_+. Applying Theorem 2, we see that u_0 is also a local $W_0^{1,p}(\Omega)$-minimizer of ψ_+, as we wished to show. □

If $u \in K_{\psi_+}$, then

$$A_p(u) + A(u) = N_{h_+}(u).$$

Acting with $(u_0 - u)^+ \in W_0^{1,p}(\Omega)$ and using (11), we show that $u_0 \le u$. Therefore

$$K_{\psi_+} \subseteq [u_0) = \big\{u \in W_0^{1,p}(\Omega) : u_0(z) \le u(z) \text{ for a.a. } z \in \Omega\big\}. \tag{30}$$

By virtue of Claim 2, $u_0 \in K_{\psi_+}$. Note that from (11) and (30) it follows that $K_{\psi_+} \subseteq K_\varphi$ and recall that by hypothesis K_φ is finite. So, as in [Aizicovici et al. 2008, proof of Proposition 29], we can find $\rho \in (0, 1)$ small such that

$$\psi_+(u_0) < \inf\,[\psi_+(u) : \|u - u_0\| = \rho] = \eta_\rho^+. \tag{31}$$

Claim 3. $\psi_+(t\hat{u}_{1,p}) \to -\infty$ *as* $t \to +\infty$.

Proof. By virtue of hypothesis H(iv), we can find $\hat{\beta}_1 \in (\hat{\beta}, 0)$ and $M_4 > \|u_0\|_\infty$ such that $f(z, x)x - pF(z, x) \le \hat{\beta}_1 x^\tau$ for a.a. $z \in \Omega$, all $x \ge M_4$. Thus

$$h_+(z, x)x - pH_+(z, x) \le \hat{\beta}_1 x^\tau + c_3 \tag{32}$$

for a.a. $z \in \Omega$, all $x \geq M_4$, some $c_3 > 0$.

(Note that $F(z,x) = H_+(z,x) - H_+(z,u_0(z)) + f(z,u_0(z))u_0(z)$ for a.a. $z \in \Omega$, all $x \geq M_4$.)

Without loss of generality, we may assume that $\tau < p$ (see H(iv)). We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{H_+(z,x)}{x^p}\right) &= \frac{h_+(z,x)x^p - px^{p-1}H_+(z,x)}{x^{2p}} \\ &= \frac{h_+(z,x)x - pH_+(z,x)}{x^{p+1}} \leq \frac{\hat{\beta}_1 x^\tau + c_3}{x^{p+1}} \quad \text{(see (32))} \\ &= \hat{\beta}_1 x^{\tau-p-1} + \frac{c_3}{x^{p+1}}.\end{aligned}$$

It follows that

$$\frac{H_+(z,x)}{x^p} - \frac{H_+(z,y)}{y^p} \leq -\frac{\hat{\beta}_1}{p-\tau}\left(\frac{1}{x^{p-\tau}} - \frac{1}{y^{p-\tau}}\right) - \frac{c_3}{p}\left(\frac{1}{x^p} - \frac{1}{y^p}\right)$$

for a.a. $z \in \Omega$, all $x \geq y \geq M_4$.

Letting $x \to +\infty$, using hypothesis H(iv) and recalling that $\tau < p$, we obtain

$$\frac{\hat{\lambda}_1(p)}{p} - \frac{H_+(z,y)}{y^p} \leq \frac{\hat{\beta}_1}{p-\tau}\frac{1}{y^{p-\tau}} + \frac{c_3}{p}\frac{1}{y^p} \quad \text{for a.a. } z \in \Omega, \text{ all } y \geq M_4,$$

or, upon multiplication by y^p and with $c_4 = c_3/p$,

$$\frac{\hat{\lambda}_1(p)}{p} y^p - H_+(z,y) \leq \frac{\hat{\beta}_1}{p-\tau} y^\tau + c_4 \quad \text{for a.a. } z \in \Omega, \text{ all } y \geq 0. \tag{33}$$

Then, for $t > 0$, we have

$$\begin{aligned}\psi_+(t\hat{u}_{1,p}) &= \frac{t^p}{p}\hat{\lambda}_1(p)\|\hat{u}_{1,p}\|_p^p + \frac{t^2}{2}\|D\hat{u}_{1,p}\|_2^2 - \int_\Omega H_+(z,t\hat{u}_{1,p})\,dz \\ &\leq \frac{\hat{\beta}_1}{p-\tau}t^\tau\|\hat{u}_{1,p}\|_\tau^\tau + \frac{t^2}{2}\|D\hat{u}_{1,p}\|_2^2 + c_4|\Omega|_N \quad \text{(see (33))}.\end{aligned} \tag{34}$$

Since $\tau > 2$ (see H(iv)) and $\hat{\beta}_1 < 0$, it follows from (34) that $\psi_+(t\hat{u}_{1,p}) \to -\infty$ as $t \to +\infty$. This proves Claim 3. $\square$

Claims 1, 3 and (31) permit the use of Theorem 1, the mountain pass theorem. So, we can find $\hat{u} \in W_0^{1,p}(\Omega)$ such that

$$\psi_+(u_0) < \eta_\rho^+ \leq \psi_+(\hat{u}) \tag{35}$$

(see (31)) and

$$\psi_+'(\hat{u}) = 0. \tag{36}$$

From (35) we see that $\hat{u} \neq u_0$, while from (36) we have $\hat{u} \in [u_0)$ (see (30)).

Therefore $\hat{u}$ is the second nontrivial positive solution of (1) (see (11)). Moreover, nonlinear regularity theory (see [Lieberman 1991]) implies that $\hat{u} \in \operatorname{int} C_+$, $u_0 \leq \hat{u}$, $u_0 \neq \hat{u}$. From the tangency principle of [Pucci and Serrin 2007, p. 35], we have

$$u_0(z) < \hat{u}(z) \quad \text{for all } z \in \Omega.$$

Let $\rho = \|\hat{u}\|_\infty$ and let $\hat{\xi} > \hat{\xi}_\rho$ ($\hat{\xi}_\rho > 0$ as postulated by hypothesis H(vi)). We set

$$h(z) = f(z, u_0(z)) + \hat{\xi} u_0(z)^{p-1} \quad \text{and} \quad \hat{h}(z) = f(z, \hat{u}(z)) + \hat{\xi} \hat{u}(z)^{p-1}.$$

Clearly, $h, \hat{h} \in L^\infty(\Omega)_+$, $h \prec \hat{h}$ (see H(vi) and recall $u_0(z) < \hat{u}(z)$ for all $z \in \Omega$). Moreover, $\hat{u} \in \operatorname{int} C_+$ and so we can use Proposition 3 and infer that $\hat{u} - u_0 \in \operatorname{int} C_+$.

Similarly, consider the truncation

$$h_-(z, x) = \begin{cases} f(z, x) & \text{if } x < v_0(z), \\ f(z, v_0(z)) & \text{if } v_0(z) \leq x. \end{cases}$$

Arguing as before, we produce a second nontrivial negative solution $\hat{v} \in -\operatorname{int} C_+$ such that $v_0 - \hat{v} \in \operatorname{int} C_+$. ☐

4. Nodal solutions

In this section we produce nodal solutions for problem (1). Under the current hypotheses H, we will produce a nodal solution, and subsequently, by strengthening the regularity on $f(z, \cdot)$ (see hypotheses $\widehat{\mathrm{H}}$ below), we will generate a second nodal solution. In this section, Morse theory is a basic tool.

Our strategy is the following. First we will show that problem (1) has extremal constant sign solutions; i.e., there is a smallest nontrivial positive solution u_+ of (1) and a biggest nontrivial negative solution v_- of (1). By truncating $f(z, \cdot)$ at $\{v_-(z), u_+(z)\}$ and using variational methods and Morse theoretic techniques, we show that problem (1) has nontrivial solutions in the order interval $[v_-, u_+]$ distinct from v_- and u_+. The extremality of v_- and u_+ implies that such solutions are necessarily nodal. The nonhomogeneity of the differential operator $u \to -\Delta_p u - \Delta u$ creates difficulties, which we have to overcome. To this end, note that hypotheses H(iii), (vii) imply that we can find $c_5 > \hat{\lambda}_1(2)$ and $c_6 > 0$ such that

$$f(z, x)x \geq c_5 x^2 - c_6 |x|^p \quad \text{for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R}.$$

This growth estimate leads to the following Dirichlet problem

$$-\Delta_p u(z) - \Delta u(z) = c_5 u(z) - c_6 |u(z)|^{p-2} u(z) \quad \text{in } \Omega, \quad u|_{\partial\Omega} = 0. \tag{37}$$

Proposition 7. *Problem* (37) *has a unique nontrivial positive solution* $u_* \in \operatorname{int} C_+$ *and, since* (37) *is odd,* $v_* = -u_* \in -\operatorname{int} C_+$ *is the unique nontrivial negative solution of* (37).

Proof. We consider the C^1-functional $\gamma_+ : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\gamma_+(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \frac{c_5}{2}\|u^+\|_2^2 + \frac{c_5}{p}\|u^+\|_p^p \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

Since $p > 2$, it is clear that γ_+ is coercive. Also, γ_+ is sequentially weakly lower semicontinuous. Therefore, we can find $u_* \in W_0^{1,p}(\Omega)$ such that

$$\gamma_+(u_*) = \inf[\gamma_+(u) : u \in W_0^{1,p}(\Omega)] = m_*^+. \tag{38}$$

Since $c_5 > \hat{\lambda}_1(2)$ and $p > 2$, for $t \in (0, 1)$ small, we have $\gamma_+(t\hat{u}_{1,2}) < 0$, which implies $\gamma_+(u_*) = m_*^+ < 0 = \gamma_+(0)$ by (38); hence $u_* \neq 0$.

From (38) we have

$$\gamma_+'(u_*) = 0 \tag{39}$$

and therefore

$$A_p(u_*) + A(u_*) = c_5 u_*^+ - c_6 (u_*^+)^{p-1}.$$

On (39) we act with $-u_*^- \in W_0^{1,p}(\Omega)$ and infer that $u_* \geq 0$, $u_* \neq 0$. Hence $A_p(u_*) + A(u_*) = c_5 u_* - c_6 u_*^{p-1}$, and so

$$-\Delta_p u_*(z) - \Delta u_*(z) = c_5 u_*(z) - c_6 u_*(z)^{p-1} \quad \text{a.e. in } \Omega, \quad u_*|_{\partial\Omega} = 0.$$

Nonlinear regularity theory (see [Ladyzhenskaya and Ural'tseva 1968; Lieberman 1991]) implies that $u_* \in C_+ \setminus \{0\}$. Moreover, from the strong maximum principle of [Pucci and Serrin 2007, p. 34], we have $u_*(z) > 0$ for all $z \in \Omega$. Then

$$\Delta_p u_*(z) + \Delta u_*(z) \leq c_6 u_*(z)^{p-1} \quad \text{a.e. in } \Omega,$$

which in view of [Pucci and Serrin 2007, p. 120] leads to

$$u_* \in \operatorname{int} C_+.$$

This establishes the existence of a nontrivial positive smooth solution of (37).

Next we show the uniqueness of $u_* \in \operatorname{int} C_+$. To this end, we consider the integral functional $\beta_+ : L^1(\Omega) \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ defined by

$$\beta_+(u) = \begin{cases} \frac{1}{p}\|Du^{1/2}\|_p^p + \frac{1}{2}\|Du^{1/2}\|_2^2 & \text{if } u \geq 0,\ u^{1/2} \in W_0^{1,p}(\Omega), \\ +\infty & \text{otherwise.} \end{cases} \tag{40}$$

Let $G_0(t) = t^p/p + t^2/2$ for all $t \geq 0$. Clearly G_0 is strictly convex and strictly increasing. We set $G(y) = G_0(\|y\|)$ for all $y \in \mathbb{R}^N$. From (40) we have

$$\beta_+(u) = \begin{cases} \int_\Omega G(Du^{1/2})\, dz & \text{if } u \geq 0,\ u^{1/2} \in W_0^{1,p}(\Omega), \\ +\infty & \text{otherwise.} \end{cases} \tag{41}$$

Let $u_1, u_2 \in \operatorname{dom}\beta_+$ and set $y_1 = u_1^{1/2}$, $y_2 = u_2^{1/2}$. Then $y_1, y_2 \in W_0^{1,p}(\Omega)$. We define

$$y_3 = (tu_1 + (1-t)u_2)^{1/2} \in W_0^{1,p}(\Omega) \quad \text{with } t \in [0, 1].$$

Then Lemma 4 of [Benguria et al. 1981] (see also [Díaz and Saá 1987, Lemma 1]) implies that

$$\|Dy_3(z)\| \le \big(t\|Dy_1(z)\|^2 + (1-t)\|Dy_2(z)\|^2\big)^{1/2} \text{ a.e. in } \Omega,$$

or again, since G_0 is increasing,

$$G_0(\|Dy_3(z)\|) \le G_0\Big(\big(t\|Dy_1(z)\|^2 + (1-t)\|Dy_2(z)\|^2\big)^{1/2}\Big) \quad \text{a.e. in } \Omega. \tag{42}$$

The right-hand side is bounded above by $tG_0(\|Dy_1(z)\|) + (1-t)G_0(\|Dy_2(z)\|)$, since $t \to G_0(t^{1/2})$ is convex. So from (42) we obtain successively

$$\begin{aligned} G(Dy_3(z)) &\le tG(Dy_1(z)) + (1-t)G(Dy_2(z)) && \text{a.e. in } \Omega,\\ G\big(D(tu_1 + (1-t)u_2)^{1/2}(z)\big) &\le tG(Du_1^{1/2}(z)) + (1-t)G(Du_2^{1/2}(z)) && \text{a.e. in } \Omega, \end{aligned}$$

and finally, using (41), the convexity of β_+.

Let $u \in W_0^{1,p}(\Omega)$ be a nontrivial positive solution of the auxiliary problem (37). From the first part of the proof we have $u \in \operatorname{int} C_+$. Therefore $u^2 \in \operatorname{dom}\beta_+$. Also, if $h \in C_0^1(\overline{\Omega})$ and $t \in (-1, 1)$ is small, then $u^2 + th \in \operatorname{dom}\beta_+$. So, the Gâteaux derivative of β_+ at u^2 in the direction h exists. The chain rule and the density of $C_0^1(\overline{\Omega})$ in $W_0^{1,p}(\Omega)$ imply

$$\beta_+'(u^2)(h) = \int_\Omega \frac{-\Delta_p u - \Delta u}{u} h\, dz \quad \text{for all } h \in W_0^{1,p}(\Omega). \tag{43}$$

Similarly, if $v \in W_0^{1,p}(\Omega)$ is another nontrivial positive solution of (37), then $v \in \operatorname{int} C_+$ and we have

$$\beta_+'(v^2)(h) = \int_\Omega \frac{-\Delta_p v - \Delta v}{v} h\, dz \quad \text{for all } h \in W_0^{1,p}(\Omega). \tag{44}$$

Since β_+ is convex, its Gâteaux derivative is monotone, and so, from (43) and (44), we have

$$\begin{aligned} 0 &\le \langle \beta_+'(u^2) - \beta_+'(v^2), u^2 - v^2 \rangle_{L^1}\\ &= \int_\Omega \Big(\frac{-\Delta_p u - \Delta u}{u} - \frac{-\Delta_p v - \Delta v}{v} \Big)(u^2 - v^2)\, dz\\ &= \int_\Omega \Big(\frac{c_5 u - c_6 u^{p-1}}{u} - \frac{c_5 v - c_6 v^{p-1}}{v} \Big)(u^2 - v^2)\, dz\\ &= c_6 \int_\Omega (v^{p-1} - u^{p-1})(u^2 - v^2)\, dz \le 0. \end{aligned}$$

Therefore $u = v$, showing that $u_* \in \operatorname{int} C_+$ is the unique nontrivial positive solution of (37).

Since (37) is odd, we conclude that $v_* = -u_* \in -\operatorname{int} C_+$ is the unique nontrivial negative solution of (37). $\square$

Having this proposition, we can now establish the existence of extremal nontrivial constant sign solutions for problem (1).

Proposition 8. *If hypotheses* H *hold, then problem* (1) *has a smallest nontrivial positive solution* $u_+ \in \operatorname{int} C_+$ *and a biggest nontrivial negative solution* $v_- \in -\operatorname{int} C_+$.

Proof. Recall that the set of nontrivial positive solutions of (1) is downward directed (i.e., if u_1, u_2 are nontrivial positive solutions of (1), then there exists a nontrivial positive solution u of (1) such that $u \le u_1$ and $u \le u_2$; see [Filippakis et al. 2009, Lemma 4.2 and Proposition 4.4]). So, in order to produce the smallest nontrivial positive solution of (1), it suffices to consider the set

$$S_+ = \{u \in W_0^{1,p}(\Omega) : u \text{ is a nontrivial solution of (1)}, u \in [0, w_+]\}.$$

From Proposition 5, we know that S_+ is nonempty and $S_+ \subseteq \operatorname{int} C_+$.

Let $\bar{u} \in S_+$ and consider the Carathéodory function

$$e_+(z, x) = \begin{cases} 0 & \text{if } x < 0, \\ c_5 x - c_6 x^{p-1} & \text{if } 0 \le x \le \bar{u}(z), \\ c_5 \bar{u}(z) - c_6 \bar{u}(z)^{p-1} & \text{if } \bar{u}(z) < x. \end{cases} \tag{45}$$

We set $E_+(z, x) = \int_0^x e_+(z, s)\, ds$ and consider the C^1-functional $\sigma_+ : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\sigma_+(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega E_+(z, u(z))\, dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

It is clear from (45) that $\sigma_+(\cdot)$ is coercive. Also, it is sequentially weakly lower semicontinuous. So, by the Weierstrass theorem, we can find $\tilde{u} \in W_0^{1,p}(\Omega)$ such that

$$\sigma_+(\tilde{u}) = \inf[\sigma_+(u) : u \in W_0^{1,p}(\Omega)]. \tag{46}$$

As before (see the proof of Proposition 7), since $c_5 > \hat{\lambda}_1(2)$ and $p > 2$, for $t \in (0, 1)$ small we have $\sigma_+(t\tilde{u}_{1,2}) < 0$, and therefore $\sigma_+(\tilde{u}) < 0 = \sigma_+(0)$; hence $\tilde{u} \ne 0$. From (46) this implies $\sigma_+'(\tilde{u}) = 0$; therefore

$$A_p(\tilde{u}) + A(\tilde{u}) = N_{e_+}(\tilde{u}). \tag{47}$$

On (47) we act with $-\tilde{u}^- \in W_0^{1,p}(\Omega)$ and obtain $\tilde{u} \ge 0$, $\tilde{u} \ne 0$ (see (45)). Also on (47) we act with $(\tilde{u}-\overline{u})^+ \in W_0^{1,p}(\Omega)$. We obtain

$$\begin{aligned}
&\langle A_p(\tilde{u}), (\tilde{u}-\overline{u})^+\rangle + \langle A(\tilde{u}), (\tilde{u}-\overline{u})^+\rangle \\
&\qquad = \int_\Omega e_+(z,\tilde{u})(\tilde{u}-\overline{u})^+\,dz = \int_\Omega (c_5\overline{u} - c_6\overline{u}^{p-1})(\tilde{u}-\overline{u})^+\,dz \quad \text{(see (45))} \\
&\qquad \le \int_\Omega f(z,\overline{u})(\tilde{u}-\overline{u})^+\,dz = \langle A_p(\overline{u}), (\tilde{u}-\overline{u})^+\rangle + \langle A(\overline{u}), (\tilde{u}-\overline{u})^+\rangle;
\end{aligned}$$

this implies

$$\int_{\{\tilde{u}>\overline{u}\}} \big(\|D\tilde{u}\|^{p-2}D\tilde{u} - \|D\overline{u}\|^{p-2}D\overline{u},\, D\tilde{u}-D\overline{u}\big)_{\mathbb{R}^N}\,dz + \|D(\tilde{u}-\overline{u})^+\|_2^2 \le 0,$$

and so $\tilde{u} \le \overline{u}$.

So, we have proved that

$$\tilde{u} \in [0,\overline{u}] = \{u \in W_0^{1,p}(\Omega) : 0 \le \tilde{u}(z) \le \overline{u}(z) \text{ a.e. in } \Omega\}, \quad \tilde{u} \ne 0.$$

From (45) and (47) it follows that

$$-\Delta_p\tilde{u}(z) - \Delta\tilde{u}(z) = c_5\tilde{u}(z) - c_6\tilde{u}(z)^{p-1} \quad \text{a.e. in } \Omega, \quad \tilde{u}|_{\partial\Omega} = 0, \quad \tilde{u} \ge 0, \quad \tilde{u} \ne 0,$$

whence $\tilde{u} = u_*$ by Proposition 7, and therefore $\tilde{u} \le \overline{u}$.

Since $\overline{u} \in S_+$ is arbitrary, we conclude that

$$u_* \le u \quad \text{for all } u \in S_+. \tag{48}$$

Now let $C \subseteq S_+$ be a chain (i.e., a totally ordered subset of S_+). Then we can find $\{u_n\}_{n\ge1} \subseteq C$ such that $\inf C = \inf_{n\ge1} u_n$; (see [Dunford and Schwartz 1958, p. 336]).

We have

$$A_p(u_n) + A(u_n) = N_f(u_n), \quad u_n \in [u_*, w_+] \quad \text{for all } n \ge 1 \tag{49}$$

by (48), so $\{u_n\}_{n\ge1} \subseteq W_0^{1,p}(\Omega)$ is bounded.

So, we may assume that

$$u_n \xrightarrow{w} u \text{ in } W_0^{1,p}(\Omega) \quad \text{and} \quad u_n \to u \text{ in } L^p(\Omega). \tag{50}$$

On (49) we act with $u_n - u \in W_0^{1,p}(\Omega)$, pass to the limit as $n \to \infty$ and use (50). Then

$$\lim_{n\to\infty} \big(\langle A_p(u_n), u_n - u\rangle + \langle A(u_n), u_n - u\rangle\big) = 0,$$

and so (reasoning as in Claim 1 in the proof of Proposition 6)

$$u_n \to u \quad \text{in } W_0^{1,p}(\Omega). \tag{51}$$

So, if in (49) we pass to the limit as $n \to \infty$ and use (51), then (48) yields

$$A_p(u) + A(u) = N_f(u), \quad u_* \le u,$$

which leads to $u \in S_+$, $u = \inf C$.

Because C is an arbitrary chain, the Kuratowski–Zorn lemma gives the existence of a minimal element $u_+ \in S_+$ of S_+. But recall that S_+ is downward directed. So, if $u \in S_+$, we can find $y \in S_+$ such that $y \le u$, $y \le u_+$. The minimality of u_+ implies that $u_+ = y$ and so $u_+ \le u$. Since $u \in S_+$ is arbitrary, we conclude that u_+ is the smallest nontrivial positive solution of (1).

Similarly, let S_- be the set of nontrivial negative solutions of (1) in $[w_-, 0]$. Then S_- is upward directed (i.e., if $v_1, v_2 \in S_-$, then we can find $v \in S_-$ such that $v_1 \le v$, $v_2 \le v$; see [Filippakis et al. 2009, Lemma 4.3]). Let $\overline{v} \in S_-$ and consider the Carathéodory function

$$e_-(z,x) = \begin{cases} c_5\overline{v}(z) - c_6|\overline{v}(z)|^{p-2}\overline{v}(z) & \text{if } x < \overline{v}(z), \\ c_5 x - c_6|x|^{p-2}x & \text{if } \overline{v}(z) \le x \le 0, \\ 0 & \text{if } 0 < x. \end{cases}$$

We set $E_-(z,x) = \int_0^x e_-(z,s)\,ds$ and consider the C^1-functional $\sigma_- : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\sigma_-(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega E_-(z,u(z))\,dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

Reasoning as above, we produce $v_- \in -\operatorname{int} C_+$, the smallest nontrivial negative solution of (1). □

To implement the strategy outlined in the beginning of this section and produce a nodal solution, we need to be able to identify the nonzero critical points of φ distinct from u_* and v_* which are in the order interval $[v_*, u_*]$. This can be done using critical groups. For this reason, we compute the critical groups of φ at the origin.

Proposition 9. *If hypotheses* H *hold, then* $C_k(\varphi, 0) = \delta_{k,d_m}\mathbb{Z}$ *for all* $k \ge 0$ *with* $d_m = \dim \bigoplus_{i=1}^m E(\hat{\lambda}_i(2))$.

Proof. Let $\mu \in (\hat{\lambda}_m(2), \hat{\lambda}_{m+1}(2))$ and consider the C^2-functional $\varphi_0 : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\varphi_0(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \frac{\mu}{2}\|u\|_2^2 \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

We consider the homotopy $h_0 : [0,1] \times W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$h_0(t,u) = t\varphi(u) + (1-t)\varphi_0(u) \quad \text{for all } (t,u) \in [0,1] \times W_0^{1,p}(\Omega).$$

Clearly $h_0(0, \cdot\,) = \varphi_0(\cdot\,)$ and $h_1(0, \cdot\,) = \varphi(\cdot\,)$.

It is easy to see that, since $p > 2$, φ_0 satisfies the C-condition. Also, reasoning as in Claim 1 in the proof of Proposition 6, via hypothesis H(iv), we show that φ satisfies the C-condition.

Suppose we can find $\{t_n\}_{n\ge 1} \subseteq [0, 1]$ and $\{u_n\}_{n\ge 1} \subseteq W_0^{1,p}(\Omega)$ such that

$$t_n \to t,\, u_n \to 0 \text{ in } W_0^{1,p}(\Omega) \quad \text{and} \quad (h_0)'_u(t_n, u_n) = 0 \quad \text{for all } n \ge 1. \tag{52}$$

From the equality in (52), we have

$$A_p(u_n) + A(u_n) = t_n N_f(u_n) + (1 - t_n)\mu u_n,$$

and therefore

$$-\Delta_p u_n(z) - \Delta u_n(z) = t_n f(z, u_n(z)) + (1 - t_n)\mu u_n(z) \text{a.e. in } \Omega, \quad u_n|_{\partial\Omega} = 0. \tag{53}$$

Since $\mu \in (\hat\lambda_m(2), \hat\lambda_{m+1}(2))$, we have

$$t f(z, x)x + (1 - t)\mu x^2 \ge c_5 x^2 - c_6|x|^p \quad \text{for a.a. } z \in \Omega, \text{ all } x \in \mathbb{R}, t \in [0, 1],$$

where $c_5 > \hat\lambda_1(2)$ and $c_6 > 0$ are as before (see (37)). Then from (53) and the proof of Proposition 8, we have $u_* \le u_n$ for all $n \ge 1$, which contradicts (52). Therefore (52) cannot happen and so the homotopy invariance of critical groups (see, for example, [Chang 2005]) implies that $C_k(h_0(0, \cdot\,), 0) = C_k(h_0(1, \cdot\,), 0)$ for all $k \ge 0$, whence

$$C_k(\varphi_0, 0) = C_k(\varphi, 0) \quad \text{for all } k \ge 0. \tag{54}$$

Note that $\varphi_0''(0) = A - \mu I$ (see [Cingolani and Vannella 2003, p. 277]) and recall that $\mu \in (\hat\lambda_m(2), \hat\lambda_{m+1}(2))$. Invoking Theorem 1.1 of [Cingolani and Vannella 2003], we have $C_k(\varphi_0, 0) = \delta_{k,d_m}\mathbb{Z}$ for all $k \ge 0$, with $d_m = \dim \bigoplus_{i=1}^m E(\hat\lambda_i(2))$. Using (54) concludes the proof. ☐

Now we have all the necessary tools to complete our strategy and produce a nodal solution.

Proposition 10. *If hypotheses* H *hold, problem* (1) *has a nodal solution* $y_0 \in C_0^1(\overline{\Omega})$ *such that*

$$u_+ - y_0 \in \operatorname{int} C_+ \quad \textit{and} \quad y_0 - v_- \in \operatorname{int} C_+.$$

Proof. Let $u_+ \in \operatorname{int} C_+$ and $v_- \in -\operatorname{int} C_+$ be the two extremal nontrivial constant sign solutions of (1) produced in Proposition 8. Using these two solutions, we introduce the following truncation of the reaction $f(z, \cdot\,)$:

$$g(z, x) = \begin{cases} f(z, v_-(z)) & \text{if } x < v_-(z), \\ f(z, x) & \text{if } v_-(z) \le x \le u_+(z), \\ f(z, u_+(z)) & \text{if } u_+(z) < x. \end{cases} \tag{55}$$

This is a Carathéodory function. We set $G(z,x) = \int_0^x g(z,s)\,ds$. Also, let $g_\pm(z,x) = g(z, \pm x^\pm)$ and $G_\pm(z,x) = \int_0^x g_\pm(z,s)\,ds$. Then we introduce the C^1-functionals $\xi^*, \xi^*_\pm : W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\xi^*(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega G(z,u(z))\,dz,$$
$$\xi^*_\pm(u) = \frac{1}{p}\|Du\|_p^p + \frac{1}{2}\|Du\|_2^2 - \int_\Omega G_\pm(z,u(z))\,dz \quad \text{for all } u \in W_0^{1,p}(\Omega).$$

Claim 1. $K_{\xi^*} \subseteq [v_-, u_+]$, $K_{\xi^*_+} = \{0, u_+\}$, $K_{\xi^*_-} = \{0, v_-\}$.

Proof. Let $u \in K_{\xi^*}$. Then we have

$$A_p(u) + A(u) = N_g(u). \tag{56}$$

On (56) we act with $(u-u_+)^+ \in W_0^{1,p}(\Omega)$ and obtain

$$\begin{aligned}\langle A_p(u), (u-u_+)^+\rangle &+ \langle A(u), (u-u_+)^+\rangle \\ &= \int_\Omega g(z,u)(u-u_+)^+\,dz = \int_\Omega f(z,u_+)(u-u_+)^+\,dz \quad \text{(see (55))} \\ &= \langle A_p(u_+), (u-u_+)^+\rangle + \langle A(u_+), (u-u_+)^+\rangle,\end{aligned}$$

so that

$$\int_{\{u>u_+\}} \big(\|Du\|^{p-2}Du - \|Du_+\|^{p-2}Du_+, Du - Du_+\big)_{\mathbb{R}^N}\,dz + \|D(u-u_+)^+\|_2^2 = 0$$

and therefore $u \le u_+$.

Similarly, acting on (56) with $(v_- - u)^+ \in W_0^{1,p}(\Omega)$, we show that $v_- \le u$. Therefore $K_{\xi^*} \subseteq [v_-, u_+]$.

In a similar fashion, we show that $K_{\xi^*_+} \subseteq [0, u_+]$. Clearly $\{0, u_+\} \subseteq K_{\xi^*_+}$. The extremality of u_+ implies that $K_{\xi^*_+} = \{0, u_+\}$. Similarly, $K_{\xi^*_-} = \{v_-, 0\}$. This proves Claim 1. □

Claim 2. *The functions $u_+ \in \operatorname{int} C_+$ and $v_- \in -\operatorname{int} C_+$ are both local minimizers of the functional ξ^*.*

Proof. It is clear from (55) that ξ^*_+ is coercive. Also, it is sequentially weakly lower semicontinuous. So, we can find $\tilde{u} \in W_0^{1,p}(\Omega)$ such that

$$\xi^*_+(\tilde{u}) = \inf[\xi^*_+(u) : u \in W_0^{1,p}(\Omega)].$$

As in the proof of Proposition 5, using hypothesis H(vii) and the fact that $2 < p$, we have $\xi^*_+(t\hat{u}_{1,2}) < 0$ for $t \in (0,1)$ small, which give $\xi^*_+(\tilde{u}) < 0 = \xi^*_+(0)$; hence $\tilde{u} \neq 0$, showing that $\tilde{u} = u_+$ by Claim 1.

But $u_+ \in \operatorname{int} C_+$ and $\xi^*|_{C_+} = \xi^*_+|_{C_+}$ (see (55)). Therefore u_+ is a local $C^1_0(\overline{\Omega})$-minimizer of ξ^*; hence it is a local $W^{1,p}_0(\Omega)$-minimizer of ξ^* (see Theorem 2). Similarly for $v_- \in -\operatorname{int} C_+$, using this time the functional ξ^*_-. $\square$

We assume that K_{ξ^*} is finite (otherwise, we already have an infinity of distinct nodal solutions). Also, without any loss of generality, we assume that $\xi^*(v_-) \le \xi^*_+(u_+)$ (the analysis is similar if the opposite inequality holds). By virtue of Claim 2, as in [Aizicovici et al. 2008, proof of Proposition 29] we can find $\rho \in (0, 1)$ small such that

$$\xi^*(v_-) \le \xi^*(u_+) < \inf[\xi^*(u) : \|u - u_+\| = \rho] = \eta^*_\rho \quad \text{and} \quad \|v_- - u_+\| > \rho. \tag{57}$$

Note that ξ^* is coercive (see (55)); hence it satisfies the C-condition. This fact and (57) permit the use of the mountain pass theorem. So, we can find $y_0 \in W^{1,p}_0(\Omega)$ such that

$$y_0 \in K_{\xi^*} \quad \text{and} \quad \eta^*_\rho \le \xi^*(y_0). \tag{58}$$

From (57), (58) and Claim 1, we have

$$y_0 \in [v_-, u_+], \quad y_0 \notin \{v_-, u_+\}. \tag{59}$$

Since y_0 is a critical point of ξ^* of mountain pass type, we have

$$C_1(\xi^*, y_0) \ne 0. \tag{60}$$

Using the homotopy invariance of critical groups, we have $C_k(\xi^*, 0) = C_k(\varphi, 0)$ for all $k \ge 0$, which gives (see Proposition 9)

$$C_k(\xi^*, 0) = \delta_{k,d_m}\mathbb{Z} \quad \text{for all } k \ge 0. \tag{61}$$

From (60), (61) and since $d_m \ge 2$, we infer that $y_0 \ne 0$. Then the extremality of u_+ and v_- and the fact that $y_0 \in [v_-, u_+]$ imply that $y_0 \in C^1_0(\overline{\Omega})$ (see [Lieberman 1991]) is a nodal solution of (1).

Using the tangency principle of [Pucci and Serrin 2007, p. 35], we have

$$v_-(z) < y_0(z) < u_+(z) \quad \text{for all } z \in \Omega. \tag{62}$$

Let $\rho = \max\{\|u_+\|, \|v_-\|\}$ and let $\xi_\rho > 0$ as postulated by hypothesis H(vi). Then, for $\tilde{\xi} > \xi_\rho$, we have

$$\begin{aligned}
&-\Delta_p y_0(z) - \Delta y_0(z) + \tilde{\xi}|y_0(z)|^{p-2}y_0(z) \\
&\qquad = f(z, y_0(z)) + \tilde{\xi}|y_0(z)|^{p-2}y_0(z) \le f(z, u_+(z)) + \tilde{\xi}u_+(z)^{p-1} \\
&\qquad = -\Delta_p u_+(z) - \Delta u_+(z) + \tilde{\xi}u_+(z)^{p-1} \quad \text{a.e. in } \Omega.
\end{aligned}$$

Setting $h(z) = f(z, y_0(z)) + \tilde{\xi}|y_0(z)|^{p-2}y_0(z)$ and $\hat{h} = f(z, u_+(z)) + \tilde{\xi}u_+(z)^{p-1}$, we see that $h, \hat{h} \in L^\infty(\Omega)$ and $h \prec \hat{h}$. Since $u_+ \in \operatorname{int} C_+$, we can apply Proposition 3 and infer that $u_+ - y_0 \in \operatorname{int} C_+$. Similarly we show that $y_0 - v_- \in \operatorname{int} C_+$. □

So, we can state the following multiplicity theorem concerning problem (1). We stress that the result is proved without assuming any differentiability on the function $x \to f(z, x)$ (see hypotheses H). In addition our multiplicity theorem provides precise sign information for all the solutions produced.

Theorem 11. *If hypotheses* H *hold, the problem* (1) *has at least five nontrivial smooth solutions*:

$$u_0, \hat{u} \in \operatorname{int} C_+ \quad \text{with } u_0 - \hat{u} \in \operatorname{int} C_+,$$
$$v_0, \hat{v} \in -\operatorname{int} C_+ \quad \text{with } v_0 - \hat{v} \in \operatorname{int} C_+,$$

and

$$y_0 \in C_0^1(\overline{\Omega}) \quad \text{nodal s.t.} \quad y_0 - v_0 \in \operatorname{int} C_+, \quad u_0 - y_0 \in \operatorname{int} C_+.$$

Next, by strengthening the regularity condition on $f(z, \cdot)$, we will be able to produce a second nodal solution.

The new hypotheses on the reaction $f(z, x)$ are the following:

Hypotheses $\widehat{\mathbf{H}}$. (i) $f : \Omega \times \mathbb{R} \to \mathbb{R}$ is a measurable function.

(ii) For a.a. $z \in \Omega$, we have $f(z, 0) = 0$ and $f(z, \cdot) \in C^1(\mathbb{R})$.

(iii) $|f_x'(z, x)| \le \alpha(z) + c|x|^{r-2}$ for a.a. $z \in \Omega$, all $x \in \mathbb{R}$, with $\alpha \in L^\infty(\Omega)_+$, $c > 0$ and $p \le r < p^*$.

(iv) If $F(z, x) = \displaystyle\int_0^x f(z, s)\, ds$, then

$$\lim_{x \to \pm\infty} \frac{pF(z, x)}{|x|^p} = \hat{\lambda}_1(p) \quad \text{uniformly for a.a. } z \in \Omega$$

and, for some $\tau > 2$,

$$\lim_{x \to \pm\infty} \frac{f(z, x)x - pF(z, x)}{|x|^\tau} \le \hat{\beta} < 0 \quad \text{uniformly for a.a. } z \in \Omega.$$

(v) There exist functions $w_\pm \in W^{1,p}(\Omega) \cap C(\overline{\Omega})$ such that

$$w_-(z) \le c_- < 0 < c_+ \le w_+(z) \quad \text{for all } z \in \overline{\Omega},$$
$$\operatorname*{ess\,sup}_{\Omega} f(\cdot, w_+(\cdot)) \le 0 \le \operatorname*{ess\,inf}_{\Omega} f(\cdot, w_-(\cdot))$$

and

$$A_p(w_-) + A(w_-) \le 0 \le A_p(w_+) + A(w_+) \quad \text{in } W^{-1,p'}(\Omega) = W_0^{1,p}(\Omega)^*$$

(where $1/p + 1/p' = 1$).

(vi) For every $\rho > 0$, there exists $\xi_\rho > 0$ such that, for a.a. $z \in \Omega$, the function $x \to f(z, x) + \xi_\rho |x|^{p-2}x$ is nondecreasing on $[-\rho, \rho]$.

(vii) There exists integer $m \geq 2$ such that

$$\hat{\lambda}_m(2) \leq f'_x(z, 0) \leq \hat{\lambda}_{m+1}(2) \quad \text{a.e. in } \Omega, \quad \hat{\lambda}_m(2) \neq f'_x(z, \cdot), \quad \hat{\lambda}_{m+1}(2) \neq f'_x(z, \cdot),$$
$$f'_x(z, 0) = \lim_{x \to 0} \frac{f(z, x)}{x} \quad \text{uniformly for a.a. } z \in \Omega.$$

In what follows, we use the notation and the functionals introduced in the proof of Proposition 10.

Proposition 12. *If hypotheses* $\widehat{\mathrm{H}}$ *hold, then problem* (1) *has a second nodal solution* $\hat{y} \in C_0^1(\overline{\Omega})$ *such that*

$$u_+ - \hat{y} \in \operatorname{int} C_+ \quad \textit{and} \quad \hat{y} - v_- \in \operatorname{int} C_+.$$

Proof. We assume that $K_{\xi_+^*}$ is finite (otherwise we already have an infinity of nodal solutions). From the proof of Proposition 10, we have

$$\{0, u_+, v_-, y_0\} \subseteq K_{\xi_+^*} \subseteq [v_-, u_+].$$

We know that $u_+ \in \operatorname{int} C_+$ and $v_- \in -\operatorname{int} C_+$ are local minimizers of the functional ξ_* (see Claim 2 in the proof of Proposition 10). So, we have

$$C_k(\xi^*, u_+) = C_k(\xi^*, v_-) = \delta_{k,0}\mathbb{Z} \quad \text{for all } k \geq 0. \tag{63}$$

Also, from (61) we have

$$C_k(\xi^*, 0) = \delta_{k,d_m}\mathbb{Z} \quad \text{for all } k \geq 0. \tag{64}$$

Moreover, since ξ^* is coercive (see (55)), we have

$$C_k(\xi^*, \infty) = \delta_{k,0}\mathbb{Z} \quad \text{for all } k \geq 0. \tag{65}$$

Claim 1. $C_k(\varphi, y_0) = C_k(\xi^*, y_0)$ *for all* $k \geq 0$.

Proof. We consider the homotopy $\tilde{h} : [0, 1] \times W_0^{1,p}(\Omega) \to \mathbb{R}$ defined by

$$\tilde{h}(t, u) = (1 - t)\xi^*(u) + t\varphi(u) \quad \text{for all } (t, u) \in [0, 1] \times W_0^{1,p}(\Omega).$$

We have $\tilde{h}(0, \cdot) = \xi^*(\cdot)$ and $\tilde{h}(1, \cdot) = \varphi(\cdot)$ and both functionals satisfy the C-condition. Let $\{t_n\}_{n\geq 1} \subseteq [0, 1]$ and $\{u_n\}_{n\geq 1} \subseteq W_0^{1,p}(\Omega)$ such that

$$t_n \to t, u_n \to y_0 \text{ in } W_0^{1,p}(\Omega) \quad \text{and} \quad (\tilde{h})'_u(t_n, u_n) = 0 \quad \text{for all } n \geq 1. \tag{66}$$

From the equation in (66), we have

$$A_p(u_n) + A(u_n) = (1 - t_n)N_g(u_n) + t_n N_f(u_n) \quad \text{for all } n \geq 1,$$

by (55). Hence

$$\begin{cases} -\Delta_p u_n(z) - \Delta u_n(z) = (1-t_n)g(z, u_n(z)) + t_n f(z, u_n(z)) & \text{a.e. in } \Omega, \\ u_n|_{\partial\Omega} = 0 & \text{for all } n \geq 1. \end{cases} \tag{67}$$

From (67) and [Lieberman 1991], we know that we can find $\gamma \in (0, 1)$ and $M > 0$ such that

$$u_n \in C_0^{1,\gamma}(\overline{\Omega}) \quad \text{and} \quad \|u_n\|_{C_0^{1,\gamma}(\overline{\Omega})} \leq M \quad \text{for all } n \geq 1. \tag{68}$$

From (68) and the compact embedding of $C_0^{1,\gamma}(\overline{\Omega})$ into $C_0^1(\overline{\Omega})$, by passing to a suitable subsequence if necessary, we may assume by (66) that

$$u_n \to y_0 \quad \text{in } C^1(\overline{\Omega}). \tag{69}$$

Since $y_0 \in \operatorname{int}_{C_0^{1,\gamma}(\overline{\Omega})}[v_-, u_+]$ (see Theorem 11), from (69) it follows that

$$u_n \in [v_-, u_+], \quad u_n \neq v_-, \quad u_n \neq u_+ \quad \text{for all } n \geq n_0;$$

this, by (55), gives $\{u_n\}_{n \geq n_0} \subseteq K_{\xi^*}$, which contradicts our hypothesis that K_{ξ^*} is finite. So, (66) cannot happen, and, from the homotopy invariance of critical groups, we have

$$C_k(\tilde{h}(0, \cdot), y_0) = C_k(\tilde{h}(1, \cdot), y_0) \quad \text{for all } k \geq 0,$$

which yields the claim. □

From (60) and Claim 1, we have

$$C_1(\varphi, y_0) \neq 0. \tag{70}$$

Claim 2. $C_k(\varphi, y_0) = \delta_{k,1}\mathbb{Z}$ *for all* $k \geq 0$.

Proof. From [Cingolani and Vannella 2007, Lemma 2.2], we know that we can find $\rho > 0$ and a C^2-function $\vartheta : V \cap \overline{B}_\rho \to \mathbb{R}$ (recall $V = H^- \oplus H^0$ (see Section 2), while $\overline{B}_\rho = \{u \in W_0^{1,p}(\Omega) : \|u\| \leq \rho\}$) such that

$$\langle \vartheta''(0)v, u\rangle = \langle \varphi''(y_0)v, u\rangle \quad \text{for all } u, v \in W_0^{1,p}(\Omega).$$

In addition $\vartheta''(0)$ is a Fredholm operator and $\ker \vartheta''(0) = H^0$. From [Cingolani and Vannella 2003, p. 286], we have

$$C_k(\varphi, y_0) = C_k(\vartheta, 0) \quad \text{for all } k \geq 0. \tag{71}$$

Then (70), (71) imply that

$$C_1(\vartheta, 0) \neq 0, \tag{72}$$

and so

$$d_- = \dim H^- \leq 1. \tag{73}$$

Let $d_0 = \dim H^0$. We consider two cases:

Case I: $d_0 = 0$. In this case, $u = 0$ is a nondegenerate critical point of ϑ with Morse index d_-. Hence $C_k(\vartheta, 0) = \delta_{k,d_-}\mathbb{Z}$ for all $k \geq 0$ (see Section 2). In view of (72) we then have $d_- = 1$, so $C_k(\varphi, y_0) = \delta_{k,1}\mathbb{Z}$ for all $k \geq 0$ (see (71)).

Case II: $d_0 > 0$. In this case, $u = 0$ is a degenerate critical point of ϑ. From (73) we see that $d_- = 1$ or $d_- = 0$.

If $d_- = 1$, then, from [Cingolani and Vannella 2003, p. 286], we have $C_k(\vartheta, 0) = \delta_{k,1}\mathbb{Z}$ for all $k \geq 0$, so $C_k(\varphi, y_0) = \delta_{k,1}\mathbb{Z}$ for all $k \geq 0$ (see (71)).

If $d_- = 0$, then, from (72) and [Cingolani and Degiovanni 2009], we have $C_k(\vartheta, 0) = \delta_{k,1}\mathbb{Z}$ for all $k \geq 0$, so $C_k(\varphi, y_0) = \delta_{k,1}\mathbb{Z}$ for all $k \geq 0$ (see (71)).

This proves Claim 2. $\square$

Claims 1 and 2 imply that

$$C_k(\xi^*, y_0) = \delta_{k,1}\mathbb{Z} \quad \text{for all } k \geq 0. \tag{74}$$

Suppose that $K_{\xi_+^*} = \{0, u_+, v_-, y_0\}$. Then, from (63), (64), (65), (74) and the Morse relation (see (4)) with $t = -1$, we have $(-1)^{d_m} = 0$, a contradiction. Therefore, we can find $\hat{y} \in K_{\xi_+^*}$, $\hat{y} \notin \{0, u_+, v_-, y_0\}$. We have $\hat{y} \in [v_-, u_+]$ (see Claim 1 in the proof of Proposition 10) and so $\hat{y}$ is nodal. Moreover, $\hat{y} \in C_0^1(\overline{\Omega})$ (nonlinear regularity) and, as we did for y_0 (see the proof of Proposition 10), we show that $\hat{y} \in \operatorname{int}_{C_0^1(\overline{\Omega})}[v_-, u_+]$. $\square$

Now we can state the second multiplicity theorem for problem (1).

Theorem 13. *If hypotheses* $\widehat{\mathrm{H}}$ *hold*, *then problem* (1) *has at least six nontrivial smooth solutions*

$$\begin{aligned} &u_0, \hat{u} \in \operatorname{int} C_+ && \text{with } \hat{u} - u_0 \in \operatorname{int} C_+,\\ &v_0, \hat{v} \in -\operatorname{int} C_+ && \text{with } v_0 - \hat{v} \in \operatorname{int} C_+ \end{aligned}$$

and $y_0, \hat{y} \in C_0^1(\overline{\Omega})$ *nodal with* $u_0 - y_0, u_0 - \hat{y} \in \operatorname{int} C_+$ *and* $y_0 - v_0, \hat{y} - v_0 \in \operatorname{int} C_+$.

Acknowledgement

The authors wish to thank a knowledgeable referee for helpful remarks and for providing additional references.

References

[Aizicovici et al. 2008] S. Aizicovici, N. S. Papageorgiou, and V. Staicu, "Degree theory for operators of monotone type and nonlinear elliptic equations with inequality constraints", *Mem. Amer. Math. Soc.* **196**:915 (2008). MR 2010c:47211 Zbl 1165.47041

[Anane 1987] A. Anane, "Simplicité et isolation de la première valeur propre du p-Laplacien avec poids", *C. R. Acad. Sci. Paris Sér. I Math.* **305**:16 (1987), 725–728. MR 89e:35124 Zbl 0633.35061

[Anane and Tsouli 1996] A. Anane and N. Tsouli, "On the second eigenvalue of the p-Laplacian", pp. 1–9 in *Nonlinear partial differential equations* (Fès, 1994), edited by A. Benkirane and J.-P. Gossez, Pitman Res. Notes Math. Ser. **343**, Longman, Harlow, 1996. MR 97k:35190 Zbl 0854.35081

[Arcoya and Ruiz 2006] D. Arcoya and D. Ruiz, "The Ambrosetti–Prodi problem for the p-Laplace operator", *Comm. Partial Differential Equations* **31**:4–6 (2006), 849–865. MR 2007h:35123 Zbl 1101.35033

[Bartsch et al. 2005] T. Bartsch, Z. Liu, and T. Weth, "Nodal solutions of a p-Laplacian equation", *Proc. London Math. Soc.* (3) **91**:1 (2005), 129–152. MR 2006a:35072 Zbl 1162.35364

[Benci et al. 2000] V. Benci, P. D'Avenia, D. Fortunato, and L. Pisani, "Solitons in several space dimensions: Derrick's problem and infinitely many solutions", *Arch. Ration. Mech. Anal.* **154**:4 (2000), 297–324. MR 2002c:35217 Zbl 0973.35161

[Benguria et al. 1981] R. Benguria, H. Brézis, and E. H. Lieb, "The Thomas–Fermi–von Weizsäcker theory of atoms and molecules", *Comm. Math. Phys.* **79**:2 (1981), 167–180. MR 83m:81114 Zbl 0478.49035

[Brézis and Nirenberg 1993] H. Brézis and L. Nirenberg, "H^1 versus C^1 local minimizers", *C. R. Acad. Sci. Paris Sér. I Math.* **317**:5 (1993), 465–472. MR 94g:49044 Zbl 0803.35029

[Chang 2005] K.-C. Chang, *Methods in nonlinear analysis*, Springer, Berlin, 2005. MR 2007b:47169 Zbl 1081.47001

[Cingolani and Degiovanni 2005] S. Cingolani and M. Degiovanni, "Nontrivial solutions for p-Laplace equations with right-hand side having p-linear growth at infinity", *Comm. Partial Differential Equations* **30**:7–9 (2005), 1191–1203. MR 2006k:35086 Zbl 1162.35367

[Cingolani and Degiovanni 2009] S. Cingolani and M. Degiovanni, "On the Poincaré–Hopf theorem for functionals defined on Banach spaces", *Adv. Nonlinear Stud.* **9**:4 (2009), 679–699. MR 2010j:58027 Zbl 1187.58016

[Cingolani and Vannella 2003] S. Cingolani and G. Vannella, "Critical groups computations on a class of Sobolev Banach spaces via Morse index", *Ann. Inst. H. Poincaré Anal. Non Linéaire* **20**:2 (2003), 271–292. MR 2004f:35033 Zbl 1023.58004

[Cingolani and Vannella 2007] S. Cingolani and G. Vannella, "Marino–Prodi perturbation type results and Morse indices of minimax critical points for a class of functionals in Banach spaces", *Ann. Mat. Pura Appl.* (4) **186**:1 (2007), 155–183. MR 2008b:58017 Zbl 1232.58006

[Derrick 1964] G. H. Derrick, "Comments on nonlinear wave equations as models for elementary particles", *J. Mathematical Phys.* **5** (1964), 1252–1254. MR 30 #4510

[Díaz and Saá 1987] J. I. Díaz and J. E. Saá, "Existence et unicité de solutions positives pour certaines équations elliptiques quasilinéaires", *C. R. Acad. Sci. Paris Sér. I Math.* **305**:12 (1987), 521–524. MR 89e:35051 Zbl 0656.35039

[Dunford and Schwartz 1958] N. Dunford and J. T. Schwartz, *Linear operators, I: General theory*, Pure Appl. Math. **7**, Wiley Interscience, New York, 1958. MR 22 #8302 Zbl 0084.10402

[Filippakis et al. 2009] M. Filippakis, A. Kristály, and N. S. Papageorgiou, "Existence of five nonzero solutions with exact sign for a p-Laplacian equation", *Discrete Contin. Dyn. Syst.* **24**:2 (2009), 405–440. MR 2010b:35122 Zbl 1169.35024

[Finn and Gilbarg 1957] R. Finn and D. Gilbarg, "Asymptotic behavior and uniqueness of plane subsonic flows", *Comm. Pure Appl. Math.* **10** (1957), 23–63. MR 19,203d Zbl 0077.18801

[García Azorero and Peral Alonso 1987] J. P. García Azorero and I. Peral Alonso, "Existence and nonuniqueness for the p-Laplacian: nonlinear eigenvalues", *Comm. Partial Differential Equations* **12**:12 (1987), 1389–1430. MR 89e:35058 Zbl 0637.35069

[García Azorero et al. 2000] J. P. García Azorero, I. Peral Alonso, and J. J. Manfredi, "Sobolev versus Hölder local minimizers and global multiplicity for some quasilinear elliptic equations", *Commun. Contemp. Math.* **2**:3 (2000), 385–404. MR 2001k:35062 Zbl 0965.35067

[Gasiński and Papageorgiou 2006] L. Gasiński and N. S. Papageorgiou, *Nonlinear analysis*, Ser. Math. Anal. Appl. **9**, Chapman & Hall/CRC, Boca Raton, FL, 2006. MR 2006e:47001 Zbl 1086.47001

[Gasiński and Papageorgiou 2012] L. Gasiński and N. S. Papageorgiou, "Multiple solutions for nonlinear coercive problems with a nonhomogeneous differential operator and a nonsmooth potential", *Set-Valued Var. Anal.* **20**:3 (2012), 417–443. MR 2949635 Zbl 06113614

[Gilbarg and Trudinger 2001] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, 2nd ed., Springer, Berlin, 2001. MR 2001k:35004 Zbl 1042.35002

[Iturriaga et al. 2010] L. Iturriaga, E. Massa, J. Sánchez, and P. Ubilla, "Positive solutions of the p-Laplacian involving a superlinear nonlinearity with zeros", *J. Differential Equations* **248**:2 (2010), 309–327. MR 2010k:35196 Zbl 1181.35117

[Ladyzhenskaya and Ural'tseva 1968] O. A. Ladyzhenskaya and N. N. Ural'tseva, *Linear and quasilinear elliptic equations*, Math. Sci. Eng. **46**, Academic Press, New York, 1968. MR 39 #5941 Zbl 0164.13002

[Lieberman 1991] G. M. Lieberman, "The natural generalization of the natural conditions of Ladyzhenskaya and Ural'tseva for elliptic equations", *Comm. Partial Differential Equations* **16**:2–3 (1991), 311–361. MR 92c:35041 Zbl 0742.35028

[Pucci and Serrin 2007] P. Pucci and J. Serrin, *The maximum principle*, Progr. Nonlinear Differential Eqns. Appl. **73**, Birkhäuser, Basel, 2007. MR 2008m:35001 Zbl 1134.35001

[Sun 2012] M. Sun, "Multiplicity of solutions for a class of the quasilinear elliptic equations at resonance", *J. Math. Anal. Appl.* **386**:2 (2012), 661–668. MR 2012h:35122 Zbl 1229.35089

[Vázquez 1984] J. L. Vázquez, "A strong maximum principle for some quasilinear elliptic equations", *Appl. Math. Optim.* **12**:3 (1984), 191–202. MR 86m:35018 Zbl 0561.35003

Received April 27, 2012. Revised January 7, 2013.

NIKOLAOS S. PAPAGEORGIOU
DEPARTMENT OF MATHEMATICS
NATIONAL TECHNICAL UNIVERSITY OF ATHENS
ZOGRAFOU CAMPUS
15780 ATHENS
GREECE

npapg@math.ntua.gr

GEORGE SMYRLIS
DEPARTMENT OF MATHEMATICS
NATIONAL TECHNICAL UNIVERSITY OF ATHENS
ZOGRAFOU CAMPUS
15780 ATHENS
GREECE

gsmyrlis@math.ntua.gr

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.455

A GEOMETRIC MODEL OF AN ARBITRARY REAL CLOSED FIELD

STANISŁAW SPODZIEJA

We give an elementary construction of any real closed field in terms of Nash function fields. We also give a characterization of any Archimedean field in terms of fields of Nash functions.

Introduction

In the study of Hilbert's 17th problem, orderings of a real field k are of importance (see [Alonso 1986; Alonso et al. 1984; Artin 1927; Artin and Schreier 1927a; 1927b; Bochnak and Efroymson 1980; Bröker 1982; Dubois 1981; Guangxing 2005; Marshall 2003; Prestel and Delzell 2001; Schwartz 1980]). By the Artin–Schreier theorem [Artin 1927; Artin and Schreier 1927a; 1927b], the study of such orderings amounts to considering real closures of k. The aim of this article is to construct a universal model of an arbitrary real closed field. To this end, we construct, in terms of Nash functions, all real closures of the rational function fields $k = \mathbb{Q}(\Lambda_T)$, where $\Lambda_T = (\Lambda_t : t \in T)$ and $T \neq \varnothing$ is a system of any number of variables. This suffices to achieve our purpose, because any real closed field R is order-preserving isomorphic to a real closure of some field $\mathbb{Q}(\Lambda_T)$ (Corollary 5.5). If $T = \varnothing$, then $\mathbb{Q}(\Lambda_T) = \mathbb{Q}$, and the above is obvious. We assume the Kuratowski–Zorn lemma, so the set T can be well-ordered, provided $T \neq \varnothing$.

L. Bröker [1982] proved in his ultrafilter theorem that there exists a one-to-one correspondence between the family of ultrafilters and the family of orderings in $\mathbb{Q}(\Lambda_T)$, or equivalently with the real closures of $\mathbb{Q}(\Lambda_T)$. We prove that there exists a one-to-one correspondence between the family of orderings in $\mathbb{Q}(\Lambda_T)$ and the family of *plain filters* (Theorem 5.2, Proposition 2.4, and Corollary 2.5). By a plain filter we mean a filter Ω of subsets of $\mathbb{R}^T$ with these properties:

(1) Any $U \in \Omega$ is a nonempty open connected semialgebraic set.

(2) For any algebraic set $V \subsetneq \mathbb{R}^T$, where $V = P^{-1}(0)$ and $P \in \mathbb{Q}[\Lambda_T]$, some connected component of $\mathbb{R}^T \setminus V$ belongs to Ω.

This research was partially supported by the program POLONIUM 2009-2010.
MSC2000: primary 14P10, 14P20; secondary 32C07.
Keywords: Nash function, semialgebraic set, real closed field, ordering.

(3) For any $U_1, U_2 \in \Omega$, there exists $U_3 \in \Omega$ such that $U_3 \subset U_1 \cap U_2$.

The correspondence between orderings and plain filters is as follows: For any ordering $\succ$ of $\mathbb{Q}(\Lambda_T)$, there exists a unique plain filter Ω such that $f \succ 0$ if and only if $f > 0$ on some $U \in \Omega$, where $>$ is the usual ordering on $\mathbb{R}$. Conversely, any plain filter Ω determines a unique ordering $\succ$ of $\mathbb{Q}(\Lambda_T)$ in this way.

The main result of this article is Theorem 5.2, where we give a construction of any real closure of $\mathbb{Q}(\Lambda_T)$ in terms of Nash functions. The main idea and motivation for the above considerations was a geometric construction of the algebraic closure of $\mathbb{C}(\Lambda_1, \dots, \Lambda_m)$ [Spodzieja 1996]. More precisely, for any plain filter Ω of open connected semialgebraic sets and any $U \in \Omega$, the ring $\mathcal{N}(U)$ of $\mathbb{Q}$-Nash functions (see Section 1) on U is a domain. In $\bigcup_{U\in\Omega} \mathcal{N}(U)$, we introduce an equivalence relation $\sim$: $(f_1 : U_1 \to \mathbb{R}) \sim (f_2 : U_2 \to \mathbb{R})$ if and only if $f_1|_{U_3} = f_2|_{U_3}$ for some $U_3 \in \Omega$. The set $\mathcal{N}_\Omega$ of equivalence classes of $\sim$ with the usual operations of addition and multiplication is a field, which is a real closure of $\mathbb{Q}(\Lambda_T)$ (see Theorem 5.2, and compare [Spodzieja 1996, Theorem 2.4 and Corollary 2.5]). One can view $\mathcal{N}_\Omega$ as the inverse limit of the étale topology $\bigcup_{U\in\Omega} \mathcal{N}(U)$ of $\mathbb{R}^T$ [Grothendieck 1967].

In Section 3, we prove that an ordering $\succ$ of $\mathbb{Q}(\Lambda_T)$ is Archimedean if and only if the set $\bigcap_{U\in\Omega} U$ is nonempty for the plain filter Ω determining $\succ$; and if that is the case, this set has exactly one point (Theorem 3.1). In Section 4, we give some examples of non-Archimedean orderings corresponding to the one in [Spodzieja 1996].

1. Preliminaries

Let $\mathbb{K}$ be the field $\mathbb{Q}$ of rational, $\mathbb{R}$ of real, or $\mathbb{C}$ of complex numbers. Let T be a nonempty set. We denote by $\Lambda_T = (\Lambda_t : t \in T)$ a system of independent variables Λ_t, by $\mathbb{K}[\Lambda_T]$ the ring of polynomials in Λ_T over $\mathbb{K}$, and by $\mathbb{K}(\Lambda_T)$ the quotient field of $\mathbb{K}[\Lambda_T]$. Note that for any $P \in \mathbb{K}(\Lambda_T)$, we have $P \in \mathbb{K}(\Lambda_{t_1}, \dots, \Lambda_{t_m})$ for some finite number of indices $t_1, \dots, t_m \in T$.

We denote by $\mathbb{K}^T$ the set of all functions $T \to \mathbb{K}$ equipped with the unique topology for which all projections $\mathbb{K}^T \ni x \mapsto x(t) \in \mathbb{K}$, $t \in T$ are continuous.

Let $\mathbb{L}$ be a subfield of $\mathbb{K}$. A subset of $\mathbb{K}^T$ is called $\mathbb{L}$-*algebraic*, or simply *algebraic* if $\mathbb{L} = \mathbb{K}$, when it is defined by a finite system of equations $P = 0$, where $P \in \mathbb{L}[\Lambda_T]$. Any $\mathbb{L}$-algebraic set in $\mathbb{K}^T$ is of the form $\{x \in \mathbb{K}^T : (x(t_1), \dots, x(t_m)) \in V\}$, where $m \in \mathbb{N}$, $t_1, \dots, t_m \in T$, and $V \subset \mathbb{K}^m$ is an $\mathbb{L}$-algebraic subset of $\mathbb{K}^m$.

If $\mathbb{L}$ is a subfield of $\mathbb{R}$, then we assume that $\mathbb{L}$ is an ordered field with order induced from $\mathbb{R}$.

Let $\mathbb{L}$ be a subfield of $\mathbb{R}$. A subset of $\mathbb{R}^T$ is called $\mathbb{L}$-*semialgebraic* when it is defined by a finite alternative of finite systems of inequalities $P > 0$ or $P \geq 0$, where $P \in \mathbb{L}[\Lambda_T]$. Analogously to the above, any $\mathbb{L}$-semialgebraic set in $\mathbb{R}^T$ is of the form

$\{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_m)) \in X\}$, where $m \in \mathbb{N}$, $t_1, \dots, t_m \in T$, and $X \subset \mathbb{R}^m$ is an $\mathbb{L}$-semialgebraic subset of $\mathbb{R}^m$. A set is called *open basic* $\mathbb{L}$*-semialgebraic* if it has the form $\{x \in \mathbb{R}^T : g_i(x) > 0,\ i = 1, \dots, n\}$, for some $n \in \mathbb{N}$ and $g_i \in \mathbb{L}[\Lambda_T]$, $i = 1, \dots, n$.

We now list some basic properties of algebraic and semialgebraic sets in infinite-dimensional real vector spaces, which follow easily from their analogues in finite-dimensional spaces [Benedetti and Risler 1990; Bochnak et al. 1987; Bochnak and Efroymson 1980; Efroymson 1974; 1976; 1981; Mostowski 1976; Prestel and Delzell 2001; Tancredi and Tognoli 2006; Tworzewski 1990].

Proposition 1.1. *Let* $\mathbb{L}$ *be a subfield of* $\mathbb{R}$ (*or* $\mathbb{K}$ *in* (a)).

(a) *The family of* $\mathbb{L}$*-algebraic sets in* $\mathbb{K}^T$ *is closed with respect to union and intersection of a finite number of sets.*

(b) *The family of* $\mathbb{L}$*-semialgebraic sets in* $\mathbb{R}^T$ *is closed with respect to complement, union, and intersection of a finite number of sets.*

(c) (Tarski–Seidenberg) *Let* $\pi_{t_1,\dots,t_m} : \mathbb{R}^T \ni x \mapsto (x(t_1), \dots, x(t_m)) \in \mathbb{R}^m$, *where* $t_1, \dots, t_m \in T$. *If* $X \subset \mathbb{R}^T$, $Y \subset \mathbb{R}^m$ *are* $\mathbb{L}$*-semialgebraic sets, then* $\pi_{t_1,\dots,t_m}(X)$ *and* $\pi^{-1}_{t_1,\dots,t_m}(Y)$ *are* $\mathbb{L}$*-semialgebraic sets, too.*

(d) *For any* $\mathbb{L}$*-semialgebraic set* $X \subset \mathbb{R}^T$, *the interior* $\operatorname{Int} X$, *closure* $\bar{X}$, *and the boundary* ∂X *are* $\mathbb{L}$*-semialgebraic sets.*

Let $\mathbb{L}$ be a subfield of $\mathbb{R}$. A function $f : U \to \mathbb{R}$, where $U \subset \mathbb{R}^T$ is an open $\mathbb{L}$-semialgebraic set, is called an $\mathbb{L}$*-Nash function* if f is analytic and there exists a nonzero polynomial $P \in \mathbb{L}[\Lambda_T, Z]$ such that $P(\lambda, f(\lambda)) = 0$ for $\lambda \in U$. In fact, f depends on a finite number of variables, so the analyticity of f is clear. The ring of $\mathbb{L}$-Nash functions in U is denoted by $\mathcal{N}^{\mathbb{L}}(U)$.

The next result follows via R. Thom's lemma (see for instance [Bochnak et al. 1987, Proposition 2.5.4 and the arguments of Theorems 2.3.6 and 2.4.4]) from the fact that any $\mathbb{L}$-semialgebraic set in a finite-dimensional vector space over $\mathbb{R}$ is the disjoint union of a finite number of $\mathbb{L}$-semialgebraic sets which are homeomorphic to Cartesian products of intervals.

Proposition 1.2. *Let* $\mathbb{L}$ *be a subfield of* $\mathbb{R}$. *Any connected component of an* $\mathbb{L}$*-semialgebraic subset of* $\mathbb{R}^T$ *is* $\mathbb{L}$*-semialgebraic.*

A function $f : U \to \mathbb{C}$, where $U \subset \mathbb{C}^T$ is an open set, is called a $\mathbb{C}$*-Nash function* if f is holomorphic and there exists a nonzero polynomial $P \in \mathbb{C}[\Lambda_T, Z]$ such that $P(\lambda, f(\lambda)) = 0$ for $\lambda \in U$. The ring of $\mathbb{C}$-Nash functions in U is denoted by $\mathcal{N}^{\mathbb{C}}(U)$.

For the basic properties of Nash functions and semialgebraic sets in finite-dimensional vector spaces, see, for instance, [Benedetti and Risler 1990; Bochnak et al. 1987; Bochnak and Efroymson 1980; Efroymson 1974; 1976; 1981; Mostowski

1976; Nash 1952; Tancredi and Tognoli 2006; Tworzewski 1990]. From these properties, we immediately obtain:

Proposition 1.3. *Let $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$, let $\mathbb{L}$ be a subfield of $\mathbb{K}$, and let $U \subset \mathbb{K}^T$ be an open connected set. Then $\mathcal{N}^{\mathbb{K}}(U)$ is a domain, provided U is semialgebraic when $\mathbb{K} = \mathbb{R}$. In particular $\mathcal{N}^{\mathbb{Q}}(U)$ is a domain.*

2. Orderings in $\mathbb{Q}(\Lambda_T)$

Let T be a nonempty set. A family Ω of subsets of $\mathbb{R}^T$ will be called a *c-filter* (connected sets filter) if it satisfies these conditions:

(i) Any $U \in \Omega$ is a nonempty open connected $\mathbb{Q}$-semialgebraic set.

(ii) For any $\mathbb{Q}$-algebraic set $V \subsetneq \mathbb{R}^T$, there exists $U \in \Omega$ such that $V \cap U = \varnothing$.

(iii) For any $U_1, U_2 \in \Omega$, there exists $U_3 \in \Omega$ such that $U_3 \subset U_1 \cap U_2$.

Proposition 2.1. *Let Ω be a c-filter of subsets of $\mathbb{R}^T$. The set $\partial\Omega := \bigcap_{U\in\Omega} \overline{U}$ has at most one point. Moreover, whenever T is a finite set, $\partial\Omega \neq \varnothing$ if and only if there exists a bounded set $U \in \Omega$.*

Proof. If $x_1, x_2 \in \partial\Omega$ with $x_1 \neq x_2$, then for some polynomial $f \in \mathbb{Q}[\Lambda_T]$, we have $f(x_1) < 0 < f(x_2)$. Hence, for some $W \in \Omega$ such that $W \cap f^{-1}(0) = \varnothing$, we have both $f(x) < 0$ and $f(x) > 0$ for some $x \in W$. This contradiction gives the first part of the assertion.

Now let $T = \{t_1, \dots, t_m\}$. Suppose that $\partial\Omega \neq \varnothing$ and each $W \in \Omega$ is an unbounded set. Take $x_0 \in \partial\Omega$, and let $f = (\Lambda_T) = \Lambda_{t_1}^2 + \cdots + \Lambda_{t_m}^2 - r$, where $r \in \mathbb{Q}$ and $r > x_0^2(t_1) + \cdots + x_0^2(t_m)$. Then $f^{-1}(0) \cap W = \varnothing$ for some $W \in \Omega$. Since W is a connected unbounded set, x_0 is not an accumulation point of W. This contradicts the choice of x_0. Now assume that some $W \in \Omega$ is bounded. Then it is easy to see that there exists a sequence of nonempty compact sets $C_1 \supset C_2 \supset \cdots$ with diameters decreasing to 0 and such that $U \cap C_n \neq \varnothing$ for all $U \in \Omega$ and $n \in \mathbb{N}$. Then there exists $x \in \bigcap_{n\in\mathbb{N}} C_n$ belonging to $\partial\Omega$. □

Let us fix a c-filter Ω and define a relation $\succ_\Omega$ in $\mathbb{Q}(\Lambda_T)$ by

$$f \succ_\Omega 0 \iff \text{there exists } U \in \Omega \text{ such that } f(x) > 0 \text{ for all } x \in U,$$

$$f \succ_\Omega g \iff f - g \succ_\Omega 0.$$

Let Ω be a family of subsets of $\mathbb{R}^T$. If an ordering $\succ$ of $\mathbb{Q}(\Lambda_T)$ satisfies $f \succ 0$ if and only if $f > 0$ on some $U \in \Omega$, we say that Ω *determines the ordering* $\succ$.

Lemma 2.2. *The relation $\succ_\Omega$ is an ordering in $\mathbb{Q}(\Lambda_T)$, or in other words, a total ordering satisfying*

$$f \succ_\Omega g \Rightarrow f + h \succ_\Omega g + h \quad \textit{and} \quad f \succ_\Omega 0,\ g \succ_\Omega 0 \Rightarrow fg \succ_\Omega 0.$$

Proof. The relation $\succ_\Omega$ is well-defined. Indeed, if $f \in \mathbb{Q}(\Lambda_T)$ and $f \neq 0$, then the union of the sets of zeros and poles of f is contained in some $\mathbb{Q}$-algebraic set $V \subsetneq \mathbb{R}^m$. Hence, by (i) and (ii), for some $U \in \Omega$, the values $f(x)$ have a fixed sign for all $x \in U$. Moreover, if for some $U_1, U_2 \in \Omega$ we have $f(x) > 0$ for $x \in U_1$ and $f(x) \leq 0$ for $x \in U_2$, then $0 < f(x) \leq 0$ for $x \in U_1 \cap U_2$, and $U_1 \cap U_2 \neq \varnothing$ by (iii). This is impossible. It is easy to see that the remaining conditions are also satisfied. □

Proposition 2.3. *Let Ω_1, Ω_2 be c-filters. If the orderings $\succ_{\Omega_1}$ and $\succ_{\Omega_2}$ are equal, then $\Omega = \{U \cup W : U \in \Omega_1,\ W \in \Omega_2\}$ is a c-filter determining the ordering $\succ_{\Omega_1}$.*

Proof. Since Ω_1 and Ω_2 are c-filters, it suffices to prove that $U \cap W \neq \varnothing$ for all $U \in \Omega_1$ and $W \in \Omega_2$. Suppose $U \cap W = \varnothing$ for some $U \in \Omega_1$ and $W \in \Omega_2$. Let $U = U_1 \cup \cdots \cup U_k \cup V$ be a decomposition of U into disjoint basic open semialgebraic sets $U_1, \ldots, U_k$ and a semialgebraic set V included in an algebraic set. By (i) and (ii), there exists $U' \in \Omega_1$ such that $U' \subset U_i$ for some $i \in \{1, \ldots, k\}$. Since $U_i = \{x \in \mathbb{R}^T : f_j(x) > 0,\ j = 1, \ldots, n\}$ for some $f_1, \ldots, f_n \in \mathbb{Q}[\Lambda_T]$, by the assumption we have $f_1, \ldots, f_n \succ_{\Omega_1} 0$, and so there exists $W_1 \in \Omega_2$ such that $f_j(x) > 0$ for all $x \in W_1$ and $j = 1, \ldots, n$. By (iii), there exists $W_2 \in \Omega_2$ such that $W_2 \subset W \cap W_1$ and $f_j(x) > 0$ for all $j = 1, \ldots, n$ and $x \in W_2$. Thus $W_2 \subset U$, which contradicts the assumption. □

Now let $\succ$ be an ordering in $\mathbb{Q}(\Lambda_T)$, and let

$$\mathcal{U}_\succ = \left\{ \bigcap_{i=1}^{n} f_i^{-1}((0, +\infty)) \subset \mathbb{R}^T : f_i \in \mathbb{Q}(\Lambda_T),\ f_i \succ 0 \text{ for } i = 1, \ldots, n,\ n \in \mathbb{N} \right\},$$

where we regard $f \in \mathbb{Q}(\Lambda_T)$ as a function $f : \mathbb{R}^T \to \mathbb{R}$. By the definition of $\mathcal{U}_\succ$ and the Tarski transfer principle (see [Tarski 1948; Seidenberg 1954]), we find that $\varnothing \notin \mathcal{U}_\succ$. Moreover, the relation $\succ$ is defined by

$$f \succ 0 \iff \text{there exists } U \in \mathcal{U}_\succ \text{ such that } f(x) > 0 \text{ for all } x \in U.$$

The sets of the family $\mathcal{U}_\succ$ may be disconnected, so $\mathcal{U}_\succ$ is not a c-filter. We will prove that the ordering $\succ$ is defined by some c-filter.

Proposition 2.4. *There exists a unique c-filter Ω with the following properties*:

(a) *For any $f \in \mathbb{Q}(\Lambda_T)$, we have $f \succ_\Omega 0$ if and only if $f \succ 0$.*

(b) *For any $U \in \Omega$, there exists a $\mathbb{Q}$-algebraic set $V \subsetneq \mathbb{R}^T$ such that U is a connected component of $\mathbb{R}^T \setminus V$.*

(c) *For any $\mathbb{Q}$-algebraic set $V \subsetneq \mathbb{R}^T$, some connected component of $\mathbb{R}^T \setminus V$ belongs to Ω.*

Proof. Let $\mathcal{F}$ be the family of all connected components of sets $U \in \mathcal{U}_\succ$.

Claim 1. *Every $U \in \mathcal{U}_{\succ}$ has a connected component U_0 such that $U_0 \cap W \neq \varnothing$ for any $W \in \mathcal{U}_{\succ}$.*

Let $U \in \mathcal{U}_{\succ}$ and let $U = U_1 \cup \cdots \cup U_n$ be the decomposition into connected components. Assume to the contrary that there exist $W_1, \ldots, W_n \in \mathcal{U}_{\succ}$ such that $U_i \cap W_i = \varnothing$ for $i = 1, \ldots, n$. Then $U \cap W_1 \cap \cdots \cap W_n = \varnothing$, which is impossible. This gives Claim 1.

Claim 2. *Each $U \in \mathcal{U}_{\succ}$ has exactly one connected component S_U that intersects every $W \in \mathcal{U}_{\succ}$.*

Let $U \in \mathcal{U}_{\succ}$, and let $U_1, \ldots, U_p$ be the connected components of U. Then

$$U = \bigcap_{l=1}^{s} \{x \in \mathbb{R}^T : g_l(x) > 0\} \tag{1}$$

for some nonzero polynomials $g_l \in \mathbb{Q}[\Lambda_T]$, with $g_l \succ 0$ for $l = 1, \ldots, s$, and

$$U_i = [f_i^{-1}(0) \cap U_i] \cup \bigcup_{j=1}^{n} \bigcap_{k=1}^{m} \{x \in \mathbb{R}^T : f_{i,j,k}(x) > 0\}, \quad i = 1, \ldots, p,$$

for some nonzero polynomials $f_i, f_{i,j,k} \in \mathbb{Q}[\Lambda_T]$. Denote by $\epsilon_{i,j,k}$ the sign of $f_{i,j,k}$ in the ordering $\succ$. Then $\epsilon_{i,j,k} \neq 0$ and $\epsilon_{i,j,k} f_{i,j,k} \succ 0$ for any i, j, k. Observe that for some $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, n\}$, we have $f_{i,j,k} \succ 0$ for $k = 1, \ldots, m$. Indeed, in the opposite case,

$$\varnothing = \bigcap_{l=1}^{s} \bigcap_{i=1}^{p} \bigcap_{j=1}^{n} \bigcap_{k=1}^{m} \{x \in \mathbb{R}^T : g_l(x) > 0,\ \epsilon_{i,j,k} f_{i,j,k}(x) > 0\} \in \mathcal{U}_{\succ},$$

which is impossible. So, for some $i_0 \in \{1, \ldots, p\}$ and $j_0 \in \{1, \ldots, n\}$,

$$U' = \bigcap_{k=1}^{m} \{x \in \mathbb{R}^T : f_{i_0, j_0, k}(x) > 0\} \in \mathcal{U}_{\succ},$$

and $U' \cap U_j = \varnothing$ for $j \neq j_0$. Hence, by Claim 1, $S_U = U_{j_0}$ is the unique connected component of U satisfying Claim 2.

Claim 3. *The family $\Omega = \{S_U : U \in \mathcal{U}_{\succ}\}$ is a c-filter.*

Since for every $\mathbb{Q}$-algebraic set $V \subset \mathbb{R}^T$ there exists $U \in \mathcal{U}_{\succ}$ such that $U \cap V = \varnothing$, we have $S_U \cap V = \varnothing$. Hence, it suffices to prove that for any $S_{U_1}, S_{U_2} \in \Omega$, there exists $S_{U_3} \in \Omega$ contained in $S_{U_1} \cap S_{U_2}$. Indeed, by the argument of Claim 2, there exist $W_1, W_2 \in \mathcal{U}_{\succ}$ such that $W_1 \subset S_{U_1}$ and $W_2 \subset S_{U_2}$. Hence, $S_{W_1 \cap W_2} \subset W_1 \cap W_2 \subset S_{U_1} \cap S_{U_2}$ and $S_{W_1 \cap W_2} \in \Omega$.

Claim 4. *The c-filter Ω defined in Claim* 3 *satisfies the assertion of Proposition 2.4.*

Part (a) is obvious.

Let $U \in \mathcal{U}_{\succ}$ be of the form (1), $f = g_1 \dots g_s$, and $V = f^{-1}(0)$. Then, by the definition of S_U, we see that S_U is a connected component of $\mathbb{R}^T \setminus V$. This gives (b).

Let $V = f^{-1}(0)$ be a $\mathbb{Q}$-algebraic subset of $\mathbb{R}^T$. Then $U = \{x \in \mathbb{R}^T : f^2(x) > 0\} = \mathbb{R}^T \setminus V \in \mathcal{U}_{\succ}$ and $S_U \in \Omega$ is a connected component of $\mathbb{R}^T \setminus V$. This gives (c) and completes the proof. □

We call the c-filter Ω defined in Proposition 2.4 the *plain filter* for the ordering $\succ$ and denote it by $\Omega_{\succ}$.

From Proposition 2.4, we immediately obtain:

Corollary 2.5. *The mapping $\succ \mapsto \Omega_{\succ}$ is a one-to-one correspondence between the set of orderings of $\mathbb{Q}(\Lambda_T)$ and the set of plain filters.*

Remark 2.6. From the ultrafilter theorem [Bröker 1982], we see that for any ultrafilter $\mathcal{F}$ of subsets of $\mathbb{R}^T$, there exists a plain filter $\Omega \subset \mathcal{F}$.

Remark 2.7. It is easy to observe that the statements of this section hold with $\mathbb{Q}$ replaced by $\mathbb{R}$.

3. Archimedean orderings in $\mathbb{Q}(\Lambda_T)$

Let $\succ$ be an ordering of $\mathbb{Q}(\Lambda_T)$. Then one can assume that T is linearly ordered by

$$t_1 \succ t_2 \iff \Lambda_{t_1} \succ \Lambda_{t_2}.$$

If $f \succ g$, then we also write $g \prec f$.

Theorem 3.1. *The following conditions are equivalent*:

(a) *The field $(\mathbb{Q}(\Lambda_T), \succ)$ is Archimedean.*

(b) *There exists $x_{\succ} \in \partial\Omega_{\succ}$ such that the set of coordinates of $x_{\succ}$ is algebraically independent over $\mathbb{Q}$.*

(c) *There exists $x_{\succ} \in \partial\Omega_{\succ}$ such that $f \succ 0$ if and only if $f(x_{\succ}) > 0$.*

(d) *There exists $x_{\succ} \in \partial\Omega_{\succ}$ such that $x_{\succ} \in U$ for any $U \in \Omega_{\succ}$.*

Proof. Assume (a). Then for any $t_1, \dots, t_n \in T$ with $t_1 \prec \dots \prec t_n$, and for the projection $\pi_{t_1,\dots,t_n} : \mathbb{R}^T \mapsto (x(t_1), \dots, x(t_n)) \in \mathbb{R}^n$, the family

$$\Omega_{t_1,\dots,t_n} = \{\pi_{t_1,\dots,t_n}(U) : U \in \Omega\} \tag{2}$$

determines an Archimedean order in $\mathbb{Q}(\Lambda_{t_1}, \dots, \Lambda_{t_n})$. Thus for some $W \in \Omega_{t_1,\dots,t_n}$, the function $f = \Lambda_{t_1}^2 + \dots + \Lambda_{t_n}^2$ is bounded on W. So the set W is bounded. Hence, by Proposition 2.1, there exists $(x_1, \dots, x_n) \in \partial\Omega_{t_1,\dots,t_n}$. Since the projections $\pi_{t_1,\dots,t_n}$ are open, it is easy to observe that, for $t_{k_1}, \dots, t_{k_j} \in \{t_1, \dots, t_n\}$ with $t_{k_1} \prec \dots \prec t_{k_j}$, we have $(x_{k_1}, \dots, x_{k_j}) \in \partial\Omega_{t_{k_1},\dots,t_{k_j}}$. Consequently, there exists $x \in \mathbb{R}^T$ such that for any $t_1, \dots, t_n \in T$ with $t_1 \prec \cdots \prec t_n$, we have $\pi_{t_1,\dots,t_n}(x) \in \partial\Omega_{t_1,\dots,t_n}$. Summing up, $x \in \partial\Omega$. The set of coordinates of x is algebraically independent over $\mathbb{Q}$: otherwise, $f(x) = 0$ for some nonzero polynomial $f \in \mathbb{Q}[\Lambda_T]$, and so f is infinitesimal. This contradicts (a) and gives (b).

Assume (b). Then any nonzero $f \in \mathbb{Q}(\Lambda_T)$ with $f \succ 0$ is defined at $x_\succ$. Moreover, $f(x_\succ) \neq 0$, so $f(x_\succ) > 0$. Conversely, assume that $f(x_\succ) > 0$. Then obviously for some connected component U of $f^{-1}(0, +\infty)$, we have $U \in \Omega_\succ$ and $f(x) > 0$ for $x \in U$. Summing up, we obtain (c).

The implication (c) $\Rightarrow$ (d) is trivial.

Now assume (d). Then we immediately obtain (b), and hence, no $f \in \mathbb{Q}(\Lambda_T)$ is infinitesimal, and the field $(\mathbb{Q}(\Lambda_T), \succ)$ is Archimedean. This gives (a) and completes the proof. □

Remark 3.2. The assertion of Theorem 3.1 also holds for every c-filter determining $\succ$ in place of the plain filter $\Omega_\succ$.

Theorem 3.1 implies:

Corollary 3.3. *Let T be a finite set. Then the set of Archimedean orderings of $\mathbb{Q}(\Lambda_T)$ is a dense subset of the space of orderings in $\mathbb{Q}(\Lambda_T)$ in the path topology* (see, for instance, [Marshall 2008]) *of the real spectrum* Sper($\mathbb{Q}[\Lambda_T]$).

4. Examples of non-Archimedean orderings

Let m be a fixed positive integer and Λ a system of m variables $\Lambda_1, \dots, \Lambda_m$.

Take any $P \in \mathbb{R}[\Lambda]$. Let $\Gamma_P \subset \mathbb{R}^m$ be a set defined by

$$\Gamma_P = \{(\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m : P(\lambda_1, \dots, \lambda_{m-1}, \lambda_m + \gamma) = 0 \text{ for some } \gamma \in [0, \infty)\}.$$

We define a polynomial $\omega(P) \in \mathbb{R}[\Lambda_1, \dots, \Lambda_{m-1}]$ (or a number $\omega(P) \in \mathbb{R}$ if $m = 1$) by $\omega(P) = 0$ for $P = 0$, and $\omega(P) = P_0$ for $P \neq 0$, where

$$P = P_0\Lambda_m^d + P_1\Lambda_m^{d-1} + \cdots + P_d$$

and $P_i \in \mathbb{R}[\Lambda_1, \dots, \Lambda_{m-1}]$ (or $P_i \in \mathbb{R}$ if $m = 1$) for $i = 0, \dots, d$ and $P_0 \neq 0$.

Let us define sets $W_P \subset \mathbb{R}^m$, for $P \in \mathbb{R}[\Lambda]$. The definition will be inductive with respect to the number of variables $\Lambda_1, \dots, \Lambda_m$. For $P \in \mathbb{R}[\Lambda]$, we put

$$W_P = \begin{cases} \mathbb{R} \setminus \Gamma_P \subset \mathbb{R} & \text{if } m = 1, \\ (\mathbb{R}^m \setminus \Gamma_P) \cap (W_{\omega(P)} \times \mathbb{R}) \subset \mathbb{R}^m & \text{if } m > 1. \end{cases} \tag{3}$$

By the Tarski–Seidenberg theorem — see Proposition 1.1(c) — the sets W_P are semialgebraic for all $P \in \mathbb{R}[\Lambda]$.

Analogously to Theorem 1.1 of [Spodzieja 1996], we prove the following proposition, which gives an example of c-filter.

Proposition 4.1. *The family* $\mathscr{W} = \{W_P : P \in \mathbb{R}[\Lambda]\}$ *satisfies these conditions*:

R_0. $W_P \subset \{\lambda \in \mathbb{R}^m : P(\lambda) \neq 0\}$.
R_1. $W_P \cap W_Q = W_{PQ}$.
R_2. *For* $P \neq 0$, W_P *is an unbounded subset of* $\mathbb{R}^m$.
R_3. *For* $P \neq 0$, W_P *is an open, connected and simply connected set.*

Moreover, one can demand that

R_4. $W_P = \mathbb{R}^m$ *for* $P = \text{const}$, $P \neq 0$.

In particular, $\mathscr{W}$ *contains the c-filter*

$$\Omega = \{W_P : P \in \mathbb{Q}[\Lambda]\}.$$

Lemma 4.2. *Let* $1 \leq i_1 < \cdots < i_m \leq n$, *and let* $P \in \mathbb{R}[\Lambda_{i_1}, \ldots, \Lambda_{i_m}]$. *Let* $Q \in \mathbb{R}[\Lambda_1, \ldots, \Lambda_n]$ *be a polynomial of the form*

$$Q(x_1, \ldots, x_n) = P(x_{i_1}, \ldots, x_{i_m}), \quad (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{4}$$

Then $W_P \subset \mathbb{R}^m$, $W_Q \subset \mathbb{R}^n$, *and*

$$W_Q \subset \{(x_1, \ldots, x_n) \in \mathbb{R}^n : (x_{i_1}, \ldots, x_{i_m}) \in W_P\}.$$

Proof. For $P = 0$ or $n = m$, the assertion is trivial. Assume that $P \neq 0$ and $n > m$. Consider the case $n = m + 1$. Then there exists $1 \leq j \leq n$ such that

$$(\Lambda_{i_1}, \ldots, \Lambda_{i_m}) = (\Lambda_1, \ldots, \Lambda_{n-j}, \Lambda_{n-j+2}, \ldots, \Lambda_n),$$

under the obvious convention for $j = 1$ and $j = n$. Denote the i-th iteration of ω by ω^i, where $\omega^0(P) = P$. Then, for $(x_1, \ldots, x_{n-i}) \in \mathbb{R}^{n-i}$,

$$\omega^i(Q)(x_1, \ldots, x_{n-i}) = \begin{cases} \omega^i(P)(x_1, \ldots, x_{n-j}, x_{n-j+2}, \ldots, x_{n-i}) & \text{if } 0 \leq i \leq j-2, \\ \omega^i(P)(x_1, \ldots, x_{n-j}) & \text{if } i = j-1, \\ \omega^{i-1}(P)(x_1, \ldots, x_{n-i}) & \text{if } j \leq i \leq n. \end{cases}$$

Hence,

$$\Gamma_{\omega^i(Q)} = \{(x_1, \ldots, x_{n-i}) \in \mathbb{R}^{n-i} : (x_1, \ldots, x_{n-j}, x_{n-j+2}, \ldots, x_{n-i}) \in \Gamma_{\omega^i(P)}\}$$

for $0 \leq i \leq j-2$, and

$$\Gamma_{\omega^{j-1}(Q)} = \{(x_1, \ldots, x_{n-j+1}) \in \mathbb{R}^{n-j+1} : (x_1, \ldots, x_{n-j}) \in \Gamma_{\omega^{j-1}(P)}\}$$

and $\Gamma_{\omega^i(Q)} = \Gamma_{\omega^{i-1}(P)}$ for $j \leq i \leq n$. In particular, $W_{\omega^i(Q)} = W_{\omega^{i-1}(P)}$ for $j \leq i \leq n$.

Summing up, by (3),

$$
\begin{aligned}
W_Q &= \bigcap_{i=0}^{n} [(\mathbb{R}^{n-i} \setminus \Gamma_{\omega^i(Q)}) \times \mathbb{R}^i] \\
&= \bigcap_{i=0}^{j-2} \{(x_1, \dots, x_n) \in \mathbb{R}^n : (x_1, \dots, x_{n-j}, x_{n-j+2}, \dots, x_{n-i}) \in \mathbb{R}^{n-i-1} \setminus \Gamma_{\omega^i(P)}\} \\
&\qquad \cap \{(x_1, \dots, x_n) \in \mathbb{R}^n : (x_1, \dots, x_{n-j}) \in \mathbb{R}^{n-j} \setminus \Gamma_{\omega^{j-1}(P)}\} \cap [W_{\omega^j(Q)} \times \mathbb{R}^j] \\
&\subset \bigcap_{i=0}^{j-2} \{(x_1, \dots, x_n) \in \mathbb{R}^n : (x_1, \dots, x_{n-j}, x_{n-j+2}, \dots, x_{n-i}) \in \mathbb{R}^{n-i-1} \setminus \Gamma_{\omega^i(P)}\} \\
&\qquad \cap [W_{\omega^{j-1}(P)} \times \mathbb{R}^j] \\
&= \{(x_1, \dots, x_n) \in \mathbb{R}^n : (x_{i_1}, \dots, x_{i_m}) \in W_P\}.
\end{aligned}
$$

This gives the assertion for $n = m + 1$. Hence, by an easy induction with respect to $n - m$, we obtain the assertion. □

Let T be a linearly ordered set and let $\succ$ be the ordering of T.

For any $t_1, \dots, t_m \in T$, $t_1 \prec \cdots \prec t_m$, we define a projection map

$$\pi_{t_1,\dots,t_m} : \mathbb{R}^T \ni x \mapsto (x(t_1), \dots, x(t_m)) \in \mathbb{R}^m.$$

Define a family Ω of semialgebraic subsets U of $\mathbb{R}^T$ by

$$U = (\pi_{t_1,\dots,t_m})^{-1}(W_P), \tag{5}$$

where $m \in \mathbb{N}$, $t_1, \dots, t_m \in T$, $t_1 \prec \cdots \prec t_m$, and $P \in \mathbb{Q}[\Lambda_{t_1}, \dots, \Lambda_{t_m}] \setminus \{0\}$.

Proposition 4.3. *The family Ω is a c-filter.*

Proof. By Proposition 4.1 (condition R_2), any $U \in \Omega$ is a nonempty set.

Let $V \subsetneq \mathbb{R}^T$ be a $\mathbb{Q}$-algebraic set, and let $P \in \mathbb{Q}[\Lambda_T] \setminus \{0\}$ be such that $V = \{x \in \mathbb{R}^T : P(x) = 0\}$. Then $P \in \mathbb{Q}[\Lambda_{t_1}, \dots, \Lambda_{t_m}] \setminus \{0\}$ for some $t_1, \dots, t_m \in T$, $t_1 \prec \cdots \prec t_m$, and $U = (\pi_{t_1,\dots,t_m})^{-1}(W_P)$. Applying Proposition 4.1 (condition R_0), we obtain that U satisfies (i).

Let $U_1, U_2 \in \Omega$. Let $t_1, \dots, t_m, u_1, \dots, u_n \in T$ satisfy $t_1 \prec \cdots \prec t_m$ and $u_1 \prec \cdots \prec u_n$, and assume moreover that for some $P \in \mathbb{Q}[\Lambda_{t_1}, \dots, \Lambda_{t_m}]$ and $Q \in \mathbb{Q}[\Lambda_{u_1}, \dots, \Lambda_{u_n}]$ we have $U_1 = (\pi_{t_1,\dots,t_m})^{-1}(W_P)$ and $U_2 = (\pi_{u_1,\dots,u_n})^{-1}(W_Q)$. Let $v_1, \dots, v_s \in T$, $v_1 \prec \cdots \prec v_s$, be such that $\{t_1, \dots, t_m\} \cup \{u_1, \dots, u_n\} \subset \{v_1, \dots, v_s\}$, and let $\overline{P}, \overline{Q} \in \mathbb{R}[\Lambda_{v_1}, \dots, \Lambda_{v_s}]$ be polynomials of the form (4) determined by P and Q, respectively. Then, by Proposition 4.1 (condition R_1) and Lemma 4.2,

$$(\pi_{v_1,\dots,v_s})^{-1}(W_{\overline{P}\,\overline{Q}}) = (\pi_{v_1,\dots,v_s})^{-1}(W_{\overline{P}}) \cap (\pi_{v_1,\dots,v_s})^{-1}(W_{\overline{Q}}) \subset U_1 \cap U_2.$$

This gives (ii).

Take any $U \in \Omega$. There exist $t_1, \dots, t_m \in T$ and $P \in \mathbb{R}[\Lambda_{t_1}, \dots, \Lambda_{t_m}] \setminus \{0\}$ such that $t_1 \prec \cdots \prec t_m$ and $U = (\pi_{t_1,\dots,t_m})^{-1}(W_P)$. By Proposition 4.1 (condition R_3), U satisfies (iii). This completes the proof. □

From the definition of the family Ω, we immediately obtain:

Corollary 4.4. *For any $t_1, t_2 \in T$, we have $t_1 \succ t_2$ if and only if $\Lambda_{t_1} \succ_\Omega \Lambda_{t_2}$.*

Let $Q \in \mathbb{Q}[\Lambda_T] \setminus \{0\}$ and let Ω_Q be a family of semialgebraic subsets U of $\mathbb{R}^T$ defined by

$$U = (\pi_{t_1,\dots,t_m})^{-1}(W_P \cap W_Q), \tag{6}$$

where $m \in \mathbb{N}$, $t_1, \dots, t_m \in T$, $t_1 \prec \cdots \prec t_m$, and $PQ \in \mathbb{Q}[\Lambda_{t_1}, \dots, \Lambda_{t_m}] \setminus \{0\}$. By Proposition 4.3, we have:

Corollary 4.5. *Ω_Q is a c-filter.*

Let $X \subset \mathbb{R}^T$ be an open semialgebraic set and let $\mathring{x} \in X$ be a point with rational coordinates. There exist $t_1, \dots, t_k \in T$, $t_1 \prec \cdots \prec t_k$, and an open semialgebraic set $Y \subset \mathbb{R}^k$ such that $X = \{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_k)) \in Y\}$. Hence, there exists $r > 0$ such that

$$B := \{x \in \mathbb{R}^T : \max_{i=1,\dots,k} |x(t_i) - \mathring{x}(t_i)| < r\} \subset X.$$

Let

$$P_0 = \Lambda_{t_1} \dots \Lambda_{t_k} (\Lambda_{t_1}^2 + \cdots + \Lambda_{t_k}^2 - 1/r^2),$$

let $U_0 = (\pi_{t_1,\dots,t_k})^{-1}(W_{P_0})$, and let $F : U_0 \to \mathbb{R}^T$ be a mapping defined by

$$F(x)(t) = \begin{cases} \mathring{x}(t) + 1/x(t) & \text{for } x \in U_0,\ t \in \{t_1, \dots, t_k\}, \\ x(t) & \text{for } x \in U_0,\ t \in T \setminus \{t_1, \dots, t_k\}. \end{cases}$$

Proposition 4.6. *$\{F(U) : U \in \Omega_{P_0}\}$ is a c-filter subset of X. In particular, for any open semialgebraic set $Y \subset \mathbb{R}^T$, there exists c-filter subsets of Y.*

Proof. By Lemma 4.2, any set $U \in \Omega_{P_0}$ is a subset of U_0. Moreover, F is an open semialgebraic mapping, so $F(U)$ is semialgebraic for $U \in \Omega_{P_0}$. Hence, $\{F(U) : U \in \Omega_{P_0}\}$ satisfies conditions (i)–(iii). □

From Proposition 4.6 and Theorem 3.1, we have that:

Corollary 4.7. *The set of c-filters defined in Proposition 4.6 is a dense subset of the space of orderings in $\mathbb{Q}(\Lambda_T)$ in the path topology of the real spectrum* Sper($\mathbb{Q}[\Lambda_T]$). *Moreover, any ordering determined by such a c-filter is not Archimedean.*

Remark 4.8. It is easy to see that the results of this section hold if we replace $\mathbb{Q}$ by $\mathbb{R}$.

5. Fields of Nash functions

Let T be a nonempty set. We denote by $\mathcal{N}(X)$ the domain of $\mathbb{Q}$-Nash functions on an open connected semialgebraic set $X \subset \mathbb{R}^T$.

Let $\succ$ be an ordering in $\mathbb{Q}(\Lambda_T)$ and let $\Omega_\succ$ be the plain filter of subsets of $\mathbb{R}^T$ determining $\succ$. Let us introduce in $\bigcup_{U\in\Omega_\succ} \mathcal{N}(U)$ a relation $\sim_\succ$ by

$$(f_1 : U_1 \to \mathbb{R}) \sim_\succ (f_2 : U_2 \to \mathbb{R}) \iff \exists_{U\in\Omega_\succ} (U \subset U_1 \cap U_2 \text{ and } f_1|_U = f_2|_U).$$

From Proposition 2.4, we immediately see that $\sim_\succ$ is an equivalence relation. The equivalence class of $\sim_\succ$ determined by $f : U \to \mathbb{R}$ is denoted by $[f]_\succ$, and the set of all such classes by $\mathcal{N}_\succ$. The set $\mathcal{N}_\succ$ is linearly ordered by

$$[f]_\succ \succ 0 \iff \exists_{U\in\Omega_\succ} (f \in \mathcal{N}(U) \text{ and } f(x) > 0 \text{ for } x \in U).$$

Proposition 5.1. *The set $\mathcal{N}_\succ$, together with the usual operations*

$$[f_1]_\succ + [f_2]_\succ = [f_1|_U + f_2|_U]_\succ\,, \quad [f_1]_\succ \cdot [f_2]_\succ = [f_1|_U f_2|_U]_\succ\,,$$

where $f_1 \in \mathcal{N}(U_1)$, $f_2 \in \mathcal{N}(U_2)$, and $U \in \Omega_\succ$, $U \subset U_1 \cap U_2$, is a real field.

Proof. Since the ring $\mathcal{N}(U)$ is a domain for any $U \in \Omega_\succ$, so is $\mathcal{N}_\succ$. We prove that any nonzero $f \in \mathcal{N}_\succ$ has an inverse in $\mathcal{N}_\succ$. Indeed, there exists $U \in \Omega_\succ$ such that $f \in \mathcal{N}(U)$. Since $f \neq 0$, the set $f^{-1}(0)$ is contained in some proper $\mathbb{Q}$-algebraic subset of $\mathbb{R}^T$. Then, by the definition of c-filter, one can assume that $f(\lambda) \neq 0$ for any $\lambda \in U$. Thus $1/f \in \mathcal{N}(U)$, so f has an inverse in $\mathcal{N}_\succ$. Summing up, $\mathcal{N}_\succ$ is a field. Since $-1 \in \mathcal{N}(U)$ is not a sum of squares in $\mathcal{N}(U)$, it follows that $-1 \in \mathcal{N}_\succ$ is not a sum of squares in $\mathcal{N}_\succ$. □

Theorem 5.2. *The field $\mathcal{N}_\succ$ is a real closure of the field* $(\mathbb{Q}(\Lambda_T), \succ)$.

Proof. Take any irreducible polynomial $P \in \mathcal{N}_\succ[Z]$ of odd degree d with respect to Z. Then there exists $U \in \Omega_\succ$ such that $P \in \mathcal{N}(U)[Z]$. Let $t_1, \dots, t_m \in T$, and let $\tilde{U} \subset \mathbb{R}^m$ be an open connected semialgebraic set such that $U = \{x \in \mathbb{R}^T : (x(t_1), \dots, x(t_m)) \in \tilde{U}\}$. By using the Hermite method (for $\tilde{U}$) we deduce that there exists a decomposition $U = U_1 \cup \cdots \cup U_k \cup V$ of U into disjoint open basic $\mathbb{Q}$-semialgebraic sets $U_1, \dots, U_k$ and a semialgebraic set V included in an algebraic set such that $P(x, Z)$ has the same number of zeroes for all $x \in U_i$ and each of these zeroes is single. By (i) and (ii) in the definition of a c-filter, there exists $U' \in \Omega_\succ$ such that $U' \subset U_i$ for some $i \in \{1, \dots, k\}$. Then there exists $k \in \mathbb{N}$, $k > 0$ such that $P(x, Z)$ has exactly k zeroes for $x \in U'$, and so there exist functions $\xi_1, \dots, \xi_k : U' \to \mathbb{R}$ with $\xi_1(x) < \cdots < \xi_k(x)$ such that $P(x, \xi_i(x)) = 0$ for $x \in U'$, $i = 1, \dots, k$. As $\xi_i(x)$ are single zeroes of $P(x, Z)$, by the Implicit Function Theorem, ξ_i is a Nash function for $i = 1, \dots, k$. As $\mathcal{N}_\succ$ is a real field

(Proposition 5.1), $\mathcal{N}_{\succ}$ is a real closed field. Since $\mathcal{N}_{\succ}$ is an algebraic extension of $\mathbb{Q}(\Lambda_T)$, by the Artin–Schreier Theorem, it is a real closure of $(\mathbb{Q}(\Lambda_T), \succ)$. □

Remark 5.3. The above results of this section also hold for an arbitrary c-filter determining $\succ$ in place of the plain filter $\Omega_{\succ}$. The results also hold if we put $\mathbb{R}$ in place of $\mathbb{Q}$.

From Theorems 3.1 and 5.2, we recover the familiar result that any Archimedean field can be embedded in $\mathbb{R}$.

Corollary 5.4. *Let $\Omega_{\succ}$ be a plain filter of subsets of $\mathbb{R}^T$ determining an Archimedean ordering $\succ$ of $\mathbb{Q}(\Lambda_T)$, and let $x_{\succ} \in \bigcap_{U\in\Omega_{\succ}} U$. Then the mapping*

$$\mathcal{N}_{\succ} \ni f \mapsto f(x_{\succ}) \in \mathbb{R}$$

is an order-preserving monomorphism.

From Theorem 5.2, we immediately obtain:

Corollary 5.5. *Let R be a real closed field with ordering $\succ$, and let T be the transcendence basis of R over $\mathbb{Q}$ whose existence is guaranteed by the Kuratowski–Zorn lemma. Assume that $T \neq \varnothing$ and let $\Lambda_T = (\Lambda_t : t \in T)$ be a system of independent variables. Then the field R is order-preserving isomorphic to a real closure of the rational functions field $\mathbb{Q}(\Lambda_T)$, i.e., to some field $\mathcal{N}_{\succ}$.*

Remark 5.6. Let $\mathbb{K}$ be an algebraically closed field of characteristic zero. Then $\mathbb{K} = R[i]$, where $i^2 = -1$, for some real closed field R. Let $T \subset R$ be the transcendence basis of $\mathbb{K}$ over $\mathbb{Q}$. Assume that $T \neq \varnothing$. Then $\mathbb{K}$ is isomorphic to an algebraic closure of $\mathbb{Q}(\Lambda_T)$. By Theorem 1.1 of [Spodzieja 1996], one can introduce a filter $\Omega_{\mathbb{C}}$ of open, connected, and simply connected semialgebraic subsets U of $\mathbb{C}^T$ satisfying conditions (i), (ii), and (iii). Then, analogously to [Spodzieja 1996], one can introduce a geometric construction of the algebraic closure of $\mathbb{Q}(\Lambda_T)$ in terms of complex Nash functions.

Acknowledgements

I would like to cordially thank Jacek Chądzyński, Michel Coste, Krzysztof Kurdyka, Tadeusz Mostowski, and Alexander Prestel for their valuable remarks and advice during the preparation of this paper.

References

[Alonso 1986] M. E. Alonso, "A note on orderings on algebraic varieties", *Pacific J. Math.* **123**:1 (1986), 1–7. MR 87k:14020 Zbl 0562.14006

[Alonso et al. 1984] M. E. Alonso, J. M. Gamboa, and J. M. Ruiz, "Ordres sur les surfaces réelles", *C. R. Acad. Sci. Paris Sér. I Math.* **298**:1 (1984), 17–19. MR 85b:32009 Zbl 0587.14011

[Artin 1927] E. Artin, "Über die Zerlegung definiter Funktionen in Quadrate", *Abh. Math. Sem. Univ. Hamburg* **5** (1927), 100–115. JFM 52.0122.01

[Artin and Schreier 1927a] E. Artin and O. Schreier, "Algebraische Konstruktion reeller Körper", *Abh. Math. Sem. Univ. Hamburg* **5** (1927), 85–99. JFM 52.0120.05

[Artin and Schreier 1927b] E. Artin and O. Schreier, "Eine Kennzeichnung der reell abgeschlossenen Körper", *Abh. Math. Sem. Univ. Hamburg* **5** (1927), 225–331. JFM 53.0144.01

[Benedetti and Risler 1990] R. Benedetti and J.-J. Risler, *Real algebraic and semi-algebraic sets*, Hermann, Paris, 1990. MR 91j:14045 Zbl 0694.14006

[Bochnak and Efroymson 1980] J. Bochnak and G. A. Efroymson, "Real algebraic geometry and the 17th Hilbert problem", *Math. Ann.* **251**:3 (1980), 213–241. MR 81k:14023 Zbl 0425.14004

[Bochnak et al. 1987] J. Bochnak, M. Coste, and M.-F. Roy, *Géométrie algébrique réelle*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **12**, Springer, Berlin, 1987. MR 90b:14030 Zbl 0633.14016

[Bröker 1982] L. Bröker, "Real spectra and distributions of signatures", pp. 249–272 in *Real algebraic geometry and quadratic forms* (Rennes, 1981), edited by J.-L. Colliot-Thélène et al., Lecture Notes in Math. **959**, Springer, Berlin, 1982. MR 84d:14014 Zbl 0501.14015

[Dubois 1981] D. W. Dubois, "Second note on Artin's solution of Hilbert's 17th problem: Order spaces", *Pacific J. Math.* **97**:2 (1981), 357–371. MR 83j:12017 Zbl 0476.14011

[Efroymson 1974] G. A. Efroymson, "A Nullstellensatz for Nash rings", *Pacific J. Math.* **54** (1974), 101–112. MR 50 #13024 Zbl 0321.14001

[Efroymson 1976] G. A. Efroymson, "Substitution in Nash functions", *Pacific J. Math.* **63**:1 (1976), 137–145. MR 53 #13211 Zbl 0335.14002

[Efroymson 1981] G. A. Efroymson, "Sums of squares in planar Nash rings", *Pacific J. Math.* **97**:1 (1981), 75–79. MR 83b:12018 Zbl 0449.32009

[Grothendieck 1967] A. Grothendieck, "Éléments de géométrie algébrique, IV: Étude locale des schémas et des morphismes de schémas IV", *Inst. Hautes Études Sci. Publ. Math.* **32** (1967), 5–361. MR 39 #220 Zbl 0153.22301

[Guangxing 2005] Z. Guangxing, "Ordered fields satisfying Pólya's theorem", *Proc. Amer. Math. Soc.* **133**:10 (2005), 2921–2926. MR 2006f:12010 Zbl 1105.12002

[Marshall 2003] M. Marshall, "$*$-orderings and $*$-valuations on algebras of finite Gelfand–Kirillov dimension", *J. Pure Appl. Algebra* **179**:3 (2003), 255–271. MR 2004a:14062 Zbl 1052.16021

[Marshall 2008] M. Marshall, *Positive polynomials and sums of squares*, Mathematical Surveys and Monographs **146**, Amer. Math. Soc., Providence, RI, 2008. MR 2009a:13044 Zbl 1169.13001

[Mostowski 1976] T. Mostowski, "Some properties of the ring of Nash functions", *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **3**:2 (1976), 245–266. MR 54 #307 Zbl 0335.14001

[Nash 1952] J. Nash, "Real algebraic manifolds", *Ann. of Math.* (2) **56** (1952), 405–421. MR 14,403b Zbl 0048.38501

[Prestel and Delzell 2001] A. Prestel and C. N. Delzell, *Positive polynomials: from Hilbert's 17th problem to real algebra*, Springer, Berlin, 2001. MR 2002k:13044 Zbl 0987.13016

[Schwartz 1980] N. Schwartz, "Archimedean lattice-ordered fields that are algebraic over their o-subfields", *Pacific J. Math.* **89**:1 (1980), 189–198. MR 82e:06018 Zbl 0442.06008

[Seidenberg 1954] A. Seidenberg, "A new decision method for elementary algebra", *Ann. of Math.* (2) **60** (1954), 365–374. MR 16,209a Zbl 0056.01804

[Spodzieja 1996] S. Spodzieja, "The field of Nash functions and factorization of polynomials", *Ann. Polon. Math.* **65**:1 (1996), 81–94. MR 97h:12006 Zbl 0909.12002

[Tancredi and Tognoli 2006] A. Tancredi and A. Tognoli, “On the products of Nash subvarieties by spheres”, *Proc. Amer. Math. Soc.* **134**:4 (2006), 983–987. MR 2006i:14062 Zbl 1093.14079

[Tarski 1948] A. Tarski, *A decision method for elementary algebra and geometry*, RAND, Santa Monica, CA, 1948. MR 10,499f Zbl 0035.00602

[Tworzewski 1990] P. Tworzewski, “Intersections of analytic sets with linear subspaces”, *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* (4) **17**:2 (1990), 227–271. MR 91j:32008 Zbl 0717.32006

Received June 11, 2012. Revised September 25, 2012.

STANISŁAW SPODZIEJA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
DEPARTMENT OF ANALYTIC FUNCTIONS AND DIFFERENTIAL EQUATIONS
UNIVERSITY OF ŁÓDŹ
S. BANACHA 22
90-238 ŁÓDŹ
POLAND

spodziej@math.uni.lodz.pl

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.471

TWISTED K-THEORY FOR THE ORBIFOLD $[*/G]$

MARIO VELÁSQUEZ, EDWARD BECERRA AND HERMES MARTINEZ

The second author dedicates this paper to Heset María

The main result of this paper establishes an explicit ring isomorphism between the twisted orbifold K-theory ${}^{\omega}K_{\mathrm{orb}}([*/G])$ and $R(D^{\omega}(G))$ for any element $\omega \in Z^3(G; S^1)$. We also study the relation between the twisted orbifold K-theories ${}^{\alpha}K_{\mathrm{orb}}(\mathscr{X})$ and ${}^{\alpha'}K_{\mathrm{orb}}(\mathscr{Y})$ of the orbifolds $\mathscr{X} = [*/G]$ and $\mathscr{Y} = [*/G']$, where G and G' are different finite groups, and $\alpha \in Z^3(G; S^1)$ and $\alpha' \in Z^3(G'; S^1)$ are different twistings. We prove that if G' is an extraspecial group with prime number p as an index and order p^n (for some fixed $n \in \mathbb{N}$), under a suitable hypothesis over the twisting α' we can obtain a twisting α on the group $(\mathbb{Z}_p)^n$ such that there exists an isomorphism between the twisted K-theories ${}^{\alpha'}K_{\mathrm{orb}}([*/G'])$ and ${}^{\alpha}K_{\mathrm{orb}}([*/(\mathbb{Z}_p)^n])$.

1. Introduction

The twisted K-theory is a successful example of the increasing flow of physical ideas into mathematics. Brought from the physical setup, the twisted orbifold K-theory has been, for the last twenty-five years, a fruitful field of ideas and development in K-theory and algebraic topology. It emerged from two sources: the consideration of the D-brane charge on a smooth manifold by Witten [1998], and the concept of discrete torsion on an orbifold by Vafa [2001]. Although for any element $\alpha \in H^3(\mathscr{X}; \mathbb{Z})$ one can associate the twisted K-theory ${}^{\alpha}K(\mathscr{X})$, its structure is simpler if the element α lies in the image of the pullback associated to the map $X \to *$. In such a case, we call this element a *discrete torsion* since we can see it as an element in the cohomology $H^3(G; \mathbb{Z})$.

On the other hand, an orbifold is a type of generalization of a smooth manifold. It is a topological space locally modeled as a quotient of a manifold by an action of a finite group. When X represents an orbifold, the twisted K-theory is more

Velásquez was partially supported by Colciencias through the grant Becas Generación del Bicentenario #494, Fundación Mazda para el Arte y la Ciencia, and Prof. Wolfgang Lück through his Leibniz Prize. We want to express our gratitude to Professor Bernardo Uribe for his important suggestions and ideas for our work.
MSC2010: primary 19L47, 19L50; secondary 20J06, 20C25.
Keywords: inverse transgression map, twisted double Drinfeld, twisted K-theory.

interesting because it is naturally related to equivariant theories if we specialize in the orbifold $\mathscr{X} = [X/G]$, where X is a smooth manifold and G is a compact Lie group acting almost freely on X. In the case of orbifolds, we have another important advantage to work with; it is the cohomological counterpart given by the Chen–Ruan cohomology on orbifolds $H^*_{CR}(\mathscr{X}; \mathbb{C})$ related to K-theory by the Chern character. The Chen–Ruan cohomology of orbifolds has an interesting nontrivial internal product which makes it an algebra. This product can be presented in the setup of the K-theory to obtain a stringy product on the K-theory of the orbifold $K_{\mathrm{orb}}(\mathscr{X})$ (see [Becerra and Uribe 2009; Adem and Ruan 2003]). If the orbifold considered has the form $[X/G]$, then the twisted orbifold K-theory can be related to the equivariant K-theory of the spaces of fixed points by the G-action on X.

On the other hand, the tensor product defines a product

$$^{\alpha}K_{\mathrm{orb}}(\mathscr{X}) \otimes {}^{\beta}K_{\mathrm{orb}}(\mathscr{X}) \to {}^{\alpha+\beta}K_{\mathrm{orb}}(\mathscr{X})$$

for any pair of elements α and β in $H^3(\mathscr{X}; \mathbb{Z})$. In fact, one can obtain a stringy product for the twisted K-theory on orbifolds by using the stringy product defined on each space of fixed points to define an explicit stringy product in each ${}^{\alpha}K(\mathscr{X})$ for any $\alpha \in H^3(\mathscr{X}; \mathbb{Z})$. Nevertheless, the crucial information to define the stringy product on the twisted K-theory of orbifolds does not lie in $H^3(\mathscr{X}; \mathbb{Z})$; instead it lies in $H^4(\mathscr{X}; \mathbb{Z})$. Given an element ϕ in $H^4(\mathscr{X}; \mathbb{Z})$, it defines an element $\theta(\phi)$ in $H^3(\wedge X; \mathbb{Z})$, where $\wedge\mathscr{X}$ is the inertia orbifold associated to $\mathscr{X}$. Hence, we can define a stringy product over the twisted K-theory orbifold ${}^{\theta(\phi)}K_{\mathrm{orb}}(\wedge\mathscr{X})$ by using a suitable structure of the inertia orbifold $\wedge X$. One such product structure is based on the map $\theta(\phi)$ called the *inverse transgression map*, which is considered to be the inverse of the classical transgression map.

For this paper, the stringy product in $K_{\mathrm{orb}}(\mathscr{X})$ has a trivial expression as we will restrict our observations to the case in which $\mathscr{X} = [\{*\}/G]$, where G is a finite group.

The main result in this paper is to present an explicit relation between the twisted Drinfeld double $D^{\omega}(G)$ and the twisted orbifold K-theory ${}^{\omega}K_{\mathrm{orb}}([*/G])$ for an element ω of discrete torsion (see Section 4). This allows us to relate the twisted orbifold K-theories ${}^{\omega}K_{\mathrm{orb}}([*/G])$ and ${}^{\omega'}K_{\mathrm{orb}}([*/G'])$ for two orbifolds $[*/G]$ and $[*/G']$ with the twistings $\omega \in H^4(G; \mathbb{Z})$ and $\omega' \in H^4(G'; \mathbb{Z})$. To obtain such a relation, we modify the stringy product defined in [Adem et al. 2007] by an element in $R_{\alpha_g}(C(g) \cap C(h))$.

2. Pushforward map in the twisted representation ring

In this section, we introduce the pushforward map. Although this map can be defined for almost complex manifolds, we will focus only on the case of homogeneous spaces G/H. To define this map, let us recall the Thom isomorphism theorem in equivariant K-theory.

Fact [Segal 1968, Proposition 3.2]. *Let X be a compact G-manifold and $p : E \to X$ be a complex G-vector bundle over X. There exists an isomorphism*

$$\phi : K_G^*(X) \to K_G^*(E, E \setminus E_0), \quad \phi([F]) := p^*(F) \otimes \lambda_{-1}(E),$$

where E_0 is the zero section and the class $\lambda_{-1}(E)$ is the Thom class associated to $[E]$.

Remark 2.1. We need to recall how to define the *normal bundle*. If M and N are G-manifolds (that means a manifold with a smooth G-action) and $f : M \to N$ is a G-embedding, we can define a (real) vector bundle τ such that $df(TM) \oplus \tau \cong TN$ (for details in this construction, consult [tom Dieck 1987]). If the map f is not a G-embedding, we can consider $f : M \to N \times D^j$ (D^j is the unitary disk in $\mathbb{R}^j$ with the trivial G-action) for sufficiently large j and by Corollary 1.10 [Wasserman 1969] we can approximate f by an immersion g_f, then we define the normal bundle for f as the normal bundle of g_f.

Now, we proceed to define the pushforward map $f_* : K_G^*(X) \to K_G^*(Y)$ for a differentiable map $f : X \to Y$ between almost complex G-manifolds by letting τ be the normal bundle associated to the map $f : X \to Y$. We define the pushforward, which will be denoted by f_*, as the composition

$$K_G^*(X) \xrightarrow{\phi} K_G^*(\tau, \tau \setminus \tau_0) \xrightarrow{j} K_G^*(Y \times D^j, (Y \times D^j) \setminus g_f(X)) \xrightarrow{i_\sharp} K_G^*(Y \times D^j) \cong K_G^*(Y),$$

where ϕ is the Thom isomorphism. The map j is given by excision, the map $i_\sharp$ is the pullback map induced by the inclusion, and the last isomorphism is induced by the natural inclusion. The pushforward map can be defined also in the twisted case (see [Carey and Wang 2008]). Consider the following diagram of inclusions:

$$\begin{tikzcd} G & K \arrow[l] \\ H \arrow[u] & H \cap K \arrow[l] \arrow[u] \end{tikzcd}$$

from which we get a diagram of surjections:

$$\begin{tikzcd} G/G & G/K \arrow[l, "i_1"'] \\ G/H \arrow[u, "j_1"] & G/(H \cap K) \arrow[l, "i_2"'] \arrow[u, "j_2"'] \end{tikzcd}$$

Using this diagram we obtain the map

$$j_{2*} \circ i_2^* : K_G^*(G/H) \to K_G^*(G/K), \quad [E] \mapsto [\lambda_{-1}(\tau_{j_2}) \otimes i_2^*(p^*(E))],$$

where τ_{j_2} is the normal bundle of j_2, and the map

$$i_1^* \circ j_{1*} : K_G^*(G/H) \to K_G^*(G/K), \quad [E] \mapsto [i_1^*(\lambda_{-1}(\tau_{j_1})) \otimes i_1^*(p^*(E))].$$

Afterwards, we compare the two maps and we conclude that the *obstruction bundle* is $\lambda_{-1}(i_1^*(\tau_{j_1})/\tau_{j_2})$. This means that

$$i_1^* \circ j_{1*}([E]) = j_{2*} \circ i_2^*([E]) \otimes \lambda_{-1}(i_1^*(\tau_{j_1})/\tau_{j_2}). \tag{2-1}$$

We consider the particular case of the groups $H = C_G(x)$ and $K = C_G(y)$, where x and y are elements in the group G and $C_G(x)$ and $C_G(y)$ denote their centralizers in G. Then by (2-1) we get an obstruction bundle which is denoted as $\gamma_{x,y}$.

3. Twisted orbifold K-theory for the orbifold $[*/G]$

The goal of this section is to consider a K-theory structure on an orbifold structure defined by the trivial action of a finite group over the space $\{*\}$. This is a particular case of a more general kind of spaces that are obtained by almost free actions of a compact Lie group over compact manifolds. These spaces naturally have an orbifold structure that sets a basis for all developments in this paper. When the manifold is one point and the group is finite, all the hypotheses in the already defined theory hold.

Let us consider the inertia orbifold $\wedge[*/G]$ for a finite group G. We define the orbifold K-theory for the orbifold $[*/G]$ as the module

$$K_{\text{orb}}([*/G]) := K(\wedge[*/G]) \cong \bigoplus_{(g)} K([*/C_G(g)]) \cong \bigoplus_{(g)} K_{C_G(g)}(*),$$

where (g) denotes the class of conjugation of the element $g \in G$ and $C_G(g)$ denotes the centralizer of the element $g \in G$. In this case, the orbifold K-theory introduced in [Adem et al. 2007] turns out to be simply $K_G(*)$, which is additively isomorphic to the group $\bigoplus_{(g)} R(C_G(g))$ (see [Adem and Ruan 2003]), where $R(C_G(g))$ denotes the Grothendieck ring associated to the semigroup of isomorphism classes of linear representations of the group $C_G(g)$, and the sum is taken over conjugacy classes. The product structure in $K_G(*)$ is defined as follows: consider the maps

$$\begin{aligned} e_1 &: C_G(g) \cap C_G(h) \times C_G(g) \cap C_G(h) \to C_G(g), & e_1(a,b) &= a, \\ e_2 &: C_G(g) \cap C_G(h) \times C_G(g) \cap C_G(h) \to C_G(h), & e_2(a,b) &= b, \\ e_{12} &: C_G(g) \cap C_G(h) \times C_G(g) \cap C_G(h) \to C_G(gh), & e_{12}(a,b) &= ab. \end{aligned}$$

Note that for any element $\tau \in G$, the map $\phi_\tau : G \to G$ defined by $\phi_\tau(g) = \tau g \tau^{-1}$ implies that $\phi_\tau \circ e_i = e_i \circ (\phi_\tau, \phi_\tau)$. Thus, the maps e_i are ϕ_τ-equivariant for any element $\tau \in C_G(g)$. Given E in $R(C_G(g))$ and F in $R(C_G(h))$, we define the product

$$E \star F := e_{12*}(e_1^*(E) \otimes e_2^*(F) \otimes \gamma_{g,h}) \in R(C_G(gh)). \tag{3-1}$$

Since the action is trivial, it follows from Theorem 2.2 in [Segal 1968] that $R(C_G(g)) = K_{C_G(g)}(*)$. Thus, the product can be seen as

$$\star : K_{C_G(g)}(*) \times K_{C_G(h)}(*) \to K_{C_G(gh)}(*)$$

in the setup of equivariant K-theory. We note that the product defined in (3-1) is analogous to the stringy and twisted stringy product defined in [Becerra and Uribe 2009] in the case in which G is an abelian group. Let α be a cocycle in $Z^3(G; S^1)$, i.e., $\alpha : G \times G \times G \to S^1$ satisfies $\alpha(a, b, c)\alpha(a, bc, d)\alpha(d, c, d) = \alpha(ab, c, d)\alpha(a, b, cd)$ for all $a, b, c, d \in G$. We proceed to define the twisted orbifold K-theory ${}^{\alpha}K_{\mathrm{orb}}([*/G])$.

For the global quotient $[X/G]$ and the element $\alpha \in Z^3(G; S^1)$, the twisted orbifold K-theory is defined as the sum

$${}^{\alpha}K_{\mathrm{orb}}([X/G]) := \bigoplus_{g \in C} {}^{\alpha_g}K_{C_G(g)}(X^g), \tag{3-2}$$

where C is a set of representatives of the conjugacy classes in G and α_g is the inverse transgression map (see below for details). In particular, if G is an abelian group, the set C is the group G. For every group H and $\beta \in Z^2(H; S^1)$ we take its associated group H_β given by the central extension

$$1 \to S^1 \to H_\beta \to H \to 1.$$

Recall that the group H_β is the set $S^1 \times H$, with the group operation defined by

$$(s_1, h_1) * (s_2, h_2) := (s_1 s_2 \beta(h_1, h_2), h_1 h_2).$$

The twisted equivariant K-theory ${}^{\beta}K_H(X)$ is defined as the class of H_β-equivariant vector bundles such that the action of the center S^1 restricts to multiplication on the fibers. In the case of the space $X = \{*\}$, the twisted equivariant K-theory ${}^{\beta}K_H(*)$ coincides with $R_\beta(H)$, the Grothendieck ring of classes of projective representations for the group H (see [Karpilovsky 1993] for a precise definition of $R_\beta(H)$).

Returning to the case of the orbifold $[*/G]$ for a finite group G, the twisted orbifold K-theory defined in (3-2) takes the form

$${}^{\alpha}K_{\mathrm{orb}}([*/G]) := \bigoplus_{g \in C} {}^{\alpha_g}K_{C_G(g)}(*) \cong \bigoplus_{g \in C} R_{\alpha_g}(C_G(g)).$$

Inverse transgression map. We review the inverse transgression map for finite groups to describe the multiplicative structure in the module ${}^{\alpha}K_{\mathrm{orb}}([*/G])$. Throughout this subsection we follow the development presented in Section 3.2 in [Becerra and Uribe 2009]. Let us recall the definition of the inverse transgression map for a global quotient $[M/G]$. For $g \in G$, consider the action of $C_G(g) \times \mathbb{Z}$ on $M^g = \{x \in M \mid gx = x\}$ given by $(h, m) \cdot x := hg^m x$ and the homomorphism

$$\psi_g : C_G(g) \times \mathbb{Z} \to G, \quad (h, m) \mapsto hg^m.$$

Thus, the inclusion $i_g : M^g \to M$ becomes a ψ_g equivariant map and induces a homomorphism

$$i_g^* : H_G^*(M; \mathbb{Z}) \to H^*_{C_G(g)\times\mathbb{Z}}(M^g; \mathbb{Z}).$$

From the isomorphisms

$$\begin{aligned} H^*_{C_G(g)\times\mathbb{Z}}(M^g; \mathbb{Z}) &\cong H^*(M^g \times_{C_G(g)} \times EC_G(g) \times B\mathbb{Z}; \mathbb{Z}) \\ &\cong H^*_{C_G(g)}(M^g; \mathbb{Z}) \otimes_{\mathbb{Z}} H^*(S^1; \mathbb{Z}), \end{aligned}$$

we have, for each k,

$$i_g^* : H_G^k(M; \mathbb{Z}) \to H^k_{C_G(g)}(M^g; \mathbb{Z}) \oplus H^{k-1}_{C_G(g)}(M^g; \mathbb{Z}).$$

Hence, we define the inverse transgression map as the map induced by projecting on the second factor

$$\tau_g : H_G^k(M; \mathbb{Z}) \to H^{k-1}_{C_G(g)}(M^g; \mathbb{Z}).$$

In the particular case that $[M/G] = [*/G]$ the definition above turns into:

Definition 3.1. For any element $\alpha \in Z^3(G; S^1)$, the inverse transgression map is defined as the map

$$\tau_g : H_G^k(*; \mathbb{Z}) \to H^{k-1}_{C_G(g)}(*; \mathbb{Z})$$

induced by τ_g on each k.

Product in the twisted case. Take $\alpha \in Z^3(G; \mathbb{Z})$. Let us consider the orbifold $[*/G]$ where G is a finite group. Now, consider the module

$$(3\text{-}3) \qquad {}^{\alpha}K_{\mathrm{orb}}([*/G]) := \bigoplus_{g \in C} R_{\alpha_g}(C_G(g)),$$

where $\alpha_g \in H^2(C(g); \mathbb{Z})$ denotes the inverse transgression map of α. The goal of this subsection is to define an associative product for this module; specifically, we show that it's possible to endow the module ${}^{\alpha}K_{\mathrm{orb}}([*/G])$ with a ring structure. For simplicity, we denote $C_G(g)$ as $C(g)$. Consider the inclusion maps of groups

$$i_g : C(g) \cap C(h) \to C(g), i_h : C(g) \cap C(h) \to C(h)$$

and

$$i_{gh} : C(g) \cap C(h) \to C(gh)$$

for $g, h \in G$. These maps induce the restriction maps

$$i_g^* : H^2(C(g); S^1) \to H^2(C(g) \cap C(h); S^1),$$
$$i_h^* : H^2(C(h); S^1) \to H^2(C(g) \cap C(h); S^1),$$

and the morphism i_{gh} induces a map

$$(i_{gh})_* : H^2(C(g) \cap C(h); S^1) \to H^2(C(gh); S^1),$$

which is the induction morphism in group cohomology (see for example [tom Dieck 1987]).

Given $E \in R_{\alpha_g}(C(g))$, we consider it as a $C(g)_{\alpha_g}$-module that restricts to multiplication on the fibers over S^1. Therefore, we get the following commutative diagram for the inclusion i_g and the identity map s on S^1:

(3-4)
$$\begin{tikzcd} 1 \arrow[r] & S^1 \arrow[r] & C(g)_{\alpha_g} \arrow[r] & C(g) \arrow[r] & 1 \\ 1 \arrow[r] & S^1 \arrow[r] \arrow[u, "s"] & (C(g) \cap C(h))_{i_g^*(\alpha_g)} \arrow[r] \arrow[u, "(s, i_g)"] & C(g) \cap C(h) \arrow[r] \arrow[u, "i_g"] & 1 \end{tikzcd}$$

This implies that any $C(g)_{\alpha_g}$-module restricts to a $(C(g) \cap C(h))_{i_g^*(\alpha_g)}$-module, denoted by $i_g^*(E)$. In particular, for any $(E, F) \in R_{\alpha_g}(C(g)) \times R_{\alpha_h}(C(h))$, we get the map

$$R_{\alpha_g}(C(g)) \times R_{\alpha_h}(C(h)) \to R_{i_g^*(\alpha_g)}(C(g) \cap C(h)) \times R_{i_h^*(\alpha_h)}(C(g) \cap C(h)),$$
$$(E, F) \mapsto (i_g^*(E), i_h^*(F)).$$

Now, from the central extensions

$$1 \to S^1 \to (C(g) \cap C(h))_{i_g^*(\alpha_g)} \to C(g) \cap C(h) \to 1,$$
$$1 \to S^1 \to (C(g) \cap C(h))_{i_h^*(\alpha_h)} \to C(g) \cap C(h) \to 1$$

induced by $i_g^*(\alpha)$ and $i_h^*(\alpha) \in H^2(C(g) \cap C(h); S^1))$, we get

$$1 \to S^1 \times S^1 \to (C(g) \cap C(h))_{i_g^*(\alpha_g)} \times (C(g) \cap C(h))_{i_h^*(\alpha_h)}$$
$$\to C(g) \cap C(h) \times C(g) \cap C(h) \to 1$$

For $E \in R_{i_g^*(\alpha_g)}(C(g))$ and $F \in R_{i_g^*(\alpha_g)}(C(g))$, the tensor product $E \otimes F$ is naturally a $(C(g) \cap C(h))_{i_g^*(\alpha_g)} \times (C(g) \cap C(h))_{i_h^*(\alpha_h)}$-module that restricts to multiplication on the fibers by elements of S^1. By considering the action restricted to the diagonal

$$\Delta(C(g)\cap C(h)) \subset C(g)\cap C(h) \times C(g)\cap C(h),$$

we get the central extension

$$1 \to S^1 \to (C(g)\cap C(h))_{i_g^*(\alpha_g)i_h^*(\alpha_h)} \to \Delta(C(g)\cap C(h)) \to 1$$

corresponding to the element

$$i_g^*(\alpha_g)i_h^*(\alpha_h) \in H^2(C(g)\cap C(h); S^1).$$

Thus, we have

$$\begin{aligned} R_{i_g^*(\alpha_g)}(C(g)\cap C(h)) \times R_{i_h^*(\alpha_h)}(C(g)\cap C(h)) &\to R_{i_g^*(\alpha_g)i_h^*(\alpha_h)}(C(g)\cap C(h)),\\ (E,F) &\mapsto E\otimes F. \end{aligned}$$

Now, since $i_g^*(\alpha_g)i_h^*(\alpha_h) = i_{gh}^*(\alpha_{gh})$ in $H^2(C(g)\cap C(h); S^1)$ since the cocycles are cohomologous (see [Adem et al. 2007, Proposition 4.3]), it follows that

$$R_{i_g^*(\alpha_g)i_h^*(\alpha_h)}(C(g)\cap C(h)) \cong R_{i_{gh}^*(\alpha_{gh})}(C(g)\cap C(h)).$$

Therefore, the induction map can be defined as

$$R_{i_{gh}^*(\alpha_{gh})}(C(g)\cap C(h)) \to R_{(i_{gh})_* i_{gh}^*(\alpha_{gh})}(C(gh)), \quad A \mapsto \mathrm{Ind}_{C(g)\cap C(h)}^{C(gh)}(A).$$

Thus, a product on the module (3-3) can be obtained from the previously described morphisms to get

$$R_{\alpha_g}(C(g)) \times R_{\alpha_h}(C(h)) \to R_{\alpha_{gh}}(C(gh)) \tag{3-5}$$

defined by

$$(E,F) \mapsto E \star_\alpha F := \mathrm{Ind}_{C(g)\cap C(h)}^{C(gh)}(i_g^*(E)\otimes i_h^*(F)\otimes \gamma_{g,h}),$$

where $\gamma_{g,h}$ is defined as the excess bundle as in Section 2.

Definition 3.2. By using the restriction notation, we define the twisted stringy product in the module ${}^\alpha K_{\mathrm{orb}}([*/G])$ as the map

$$\begin{aligned} &R_{\alpha_g}(C(g)) \times R_{\alpha_h}(C(h)) \to R_{\alpha_g\alpha_h}(C(gh))\\ &\qquad (E,F) \mapsto I_{C(g)\cap C(h)}^{C(gh)}\big(\mathrm{Res}_{C(g)\cap C(h)}^{C(g)}(E)\otimes \mathrm{Res}_{C(g)\cap C(h)}^{C(h)}(F)\otimes \gamma_{g,h}\big), \end{aligned}$$

where $\mathrm{Res}_{C(g)\cap C(h)}^{C(g)}$ denotes the restriction of α-twisted representations of $C(g)$ to $i_g(\alpha)$-representations of $C(g)\cap C(h)$ (and likewise for h).

4. Twisted orbifold K-theory and the algebra $D^\omega(G)$

The goal of this section is to give an introduction of the *twisted Drinfeld double* $D^\omega(G)$ and to show an explicit relation with the twisted orbifold K-theory. Our main reference is [Witherspoon 1996]. Let us recall the definition and the main properties of the twisted Drinfeld double to clarify the nature of this structure and its representations. From a different point of view, we can also obtain some properties of the stringy product defined on the sections above, using the properties of the representations of the twisted Drinfeld double. Namely, the Grothendieck ring of these representations is isomorphic to the twisted orbifold K-theory with the structure induced by the stringy product, which will be proven on page 481. Because of the associativity of the tensor product of the $D^\omega(G)$-modules, this yields a proof of the associativity of the stringy product defined above; see Corollary 4.3.

Let G be a finite group and k an algebraically closed field. Let ω be an element in $Z^3(G, k^*)$, that is, a function $\omega : G \times G \times G \to k^*$ such that

$$\omega(a, b, c)\,\omega(a, bc, d)\,\omega(d, c, d) = \omega(ab, c, d)\,\omega(a, b, cd)$$

for all $a, b, c, d \in G$. We define the quasitriangular quasi-Hopf algebra $D^\omega(G)$ as the vector space $(kG)^* \otimes (kG)$, where $(kG)^*$ denotes the dual of the algebra kG (see [Drinfel'd 1987]) and the algebra structure in $D^\omega(G)$ is given as follows: consider the canonical basis $\{\delta_g \otimes \bar{x}\}_{g,x\in G}$ of $D^\omega(G)$, where δ_g is the function such that $\delta_g(h) = 1$ if $h = g$ and 0 otherwise. We denote $\delta_g \otimes \bar{x}$ by $\delta_g\bar{x}$. Now, we define the product of elements in the basis by

$$(\delta_g\bar{x})(\delta_h\bar{y}) = \omega_g(x, y)\,\delta_g\delta_{xhx^{-1}}\overline{xy}, \tag{4-1}$$

where ω_g is the image of ω via the inverse transgression map of the element $g \in G$ as in Definition 3.1. The multiplicative identity for this product is the element $1_{D^\omega(G)} = \bigoplus_{g\in G} \delta_g\bar{1}$. Now, we use the notation δ_g for the element $\delta_g\bar{1}$. The coproduct $\Delta : D^\omega(G) \to D^\omega(G) \otimes D^\omega(G)$ in the algebra $D^\omega(G)$ is defined by the map

$$\Delta(\delta_g\bar{x}) = \bigoplus_{h\in G} \gamma_x(h, h^{-1}g)(\delta_h\bar{x}) \otimes (\delta_{h^{-1}g}\bar{x}), \tag{4-2}$$

where

$$\gamma_x(h, l) = \frac{\omega(h, l, x)\,\omega(x, x^{-1}hx, x^{-1}lx)}{\omega(h, x, x^{-1}lx)}.$$

The algebra $D^\omega(G)$ endowed with these operations is usually called the twisted Drinfeld double.

Representations of $D^\omega(G)$. Let U, V be modules over the algebra $D^\omega(G)$. Consider the tensor product $U \otimes V$ as a $D^\omega(G)$-module endowed with the action from $D^\omega(G)$, induced by the coproduct Δ. Note that the field k can be considered as a trivial $D^\omega(G)$-module, which is the multiplicative identity for the tensor product of $D^\omega(G)$-modules. In particular, for $k=\mathbb{C}$, we define the ring of representations $R(D^\omega(G))$ of $D^\omega(G)$ as the $\mathbb{C}$-algebra generated by the set of isomorphism classes of $D^\omega(G)$-modules with the direct sum of modules as the sum operation and the tensor as the product operation. We define the ideal $R_0(D^\omega(G))$ generated by all combinations $[U]-[U']-[U'']$ (brackets denoting the isomorphism class) where $0\to U'\to U\to U''\to 0$ is a short exact sequence of $D^\omega(G)$-modules. Now, we define the Grothendieck ring $R(D^\omega(G))$ as the quotient between $\mathrm{Rep}(D^\omega(G))$ and the ideal $R_0(D^\omega(G))$.

The algebra $D^\omega(G)$ is quasitriangular with

$$A=\bigoplus_{g,h\in G}\delta_g\bar{1}\otimes\delta_h\bar{g}\quad\text{and}\quad A^{-1}=\bigoplus_{g,h\in G}\omega_{ghg^{-1}}(g,g^{-1})^{-1}\delta_g\bar{1}\otimes\delta_h\overline{g^{-1}}.$$

Thus, $A\Delta(a)A^{-1}=\sigma(\Delta(a))$ for all $a\in D^\omega(G)$, where σ is the automorphism that exchanges the images in the coproduct. Therefore, if U and V are $D^\omega(G)$-modules, this equation implies that $U\otimes V$ and $V\otimes U$ are isomorphic as $D^\omega(G)$-modules; that is, the algebra $R(D^\omega(G))$ is commutative. Now, assume that $\beta: G\times G\to\mathbb{C}^*$ is a cochain with coboundary

$$\delta\beta(a,b,c)=\beta(b,c)\beta(a,bc)\beta(ab,c)^{-1}\beta(a,b)^{-1}.$$

Then, the algebra $D^{\omega\delta\beta}(G)$ is isomorphic to $D^\omega(G)$ given through the map

$$\upsilon(\delta_g\bar{x})=\frac{\beta(g,x)}{\beta(x,xgx^{-1})}\delta_g\bar{x}.$$

In particular, we get the isomorphism

$$\upsilon^*: R(D^{\omega\delta\beta}(G))\xrightarrow{\cong}R(D^\omega(G)).$$

Next, we consider the following theorem (compare [Willerton 2008, Theorem 19]):

Theorem 4.1. *The ring $R(D^\omega(G))$ is additively isomorphic to the ring*

$$\bigoplus_{(g)\subset G}R_{\omega_g}(C(g)),$$

where (g) denotes the conjugacy class of $g\in G$.

Proof. For all $x\in G$, we take the subspaces

$$S^\omega(x):=\bigoplus_{g\in C(x)}\mathbb{C}\delta_x\bar{g}\quad\text{and}\quad D^\omega(x):=\bigoplus_{g\in G}\mathbb{C}\delta_x\bar{g}$$

of $D^\omega(G)$. Then $S^\omega(x)$ is a subalgebra of $D^\omega(G)$ with identity element $\delta_x\bar{1}$ such that, from the product defined in $D^\omega(G)$, it follows that $S^\omega(x)\cong R_{\omega_x}C(x)$ where

$R_{\omega_x}C(x)$ is defined in [Karpilovsky 1993]. Given $(g) \subset G$, consider

$$D^{\omega}((g)) := \bigoplus_{h \in (g)} D^{\omega}(h).$$

Note that $D^{\omega}(G) \cong \bigoplus_{(g)\subset G} D^{\omega}((g))$ (additively). For an element h in a fixed conjugacy class (g), take a $S^{\omega}(h)$-module (i.e., a $R_{\omega_h}C(h)$-module) U, and define the map

$$U \mapsto U \otimes_{S^{\omega}(h)} D^{\omega}(h),$$

whose image is a $D^{\omega}((g))$-module if we take the action of $D^{\omega}((g))$ on it as right multiplication in the second factor. On the other hand, for a $D^{\omega}((g))$-module V, we define the map

$$V \mapsto V\delta_h \bar{1},$$

whose image is a $R_{\omega_h}C(h)$-module. Thus, there is an equivalence between $R_{\omega_h}C(h)$-modules and $D^{\omega}((g))$-modules. Therefore, from [Karpilovsky 1993, Theorem I.3.2], we have $R_{\omega_h}(C(h)) \cong R(D^{\omega}((g)))$ for any $h \in (g)$, and the theorem follows. □

From [Dijkgraaf et al. 1991], we get that it is possible to explicitly describe the morphism using the *induction DPR* which is defined on each $R_{\alpha}(C(g))$ for $g \in G$. Namely, let (ρ, V) be a twisted representation of the group $C(g)$ and define the representation $\psi((\rho, V)) := (\pi_{\rho}, A)$ of $D^{\omega}(G)$ as given by

$$A := \mathrm{Ind}_{C(g)}^{G}(V), \qquad \pi_{\rho} := \pi_{\rho}(\delta_k \bar{x})x_j \otimes v = \delta_k \delta_{x_s g x_s^{-1}} \frac{\omega_k(x, x_j)}{\omega_k(x_s, r)} x_s \otimes \rho(r)v, \tag{4-3}$$

where x_j is a representative of a class in $G/C(g)$, $r \in C(g)$ and the element x_s is a representative of a class in $G/C(g)$, such that $xx_j = x_s r$.

Relation between $R(D^{\omega}(G))$ and the twisted K-theory of the orbifold $[*/G]$. Let us consider an element $\omega \in Z^3(G; S^1)$. By (3-3) the twisted orbifold K-theory of the orbifold $[*/G]$ is the ring

$${}^{\omega}K_{\mathrm{orb}}([*/G]) = \bigoplus_{(g)\subset G} {}^{\omega_g}K_{C(g)}(*) \cong \bigoplus_{(g)\subset G} R_{\omega_g}(C(g)).$$

By Theorem 4.1, there exists an additive isomorphism between $R(D^{\omega}(G))$ and the twisted orbifold K-theory ${}^{\omega}K_{\mathrm{orb}}([*/G])$. We will show that if we endow this ring with the twisted product $\star_{\alpha}$, then the additive isomorphism is in fact a ring isomorphism. The DPR induction is defined as $(I_{C(g)}^{G}(E), \rho_{\pi})$ where (E, π) is an element in $R_{\omega_g}(C(g))$. Let us consider two elements E and F in $R_{\omega_g}(C(g))$ and $R_{\omega_h}(C(h))$ respectively. The tensor product of the DPR-induction of these elements can be related to the twisted product $\star$ via the Frobenius reciprocity as follows:

$$\begin{aligned}
I^G_{C(g)}(E)\otimes I^G_{C(h)}(F) &\cong I^G_{C(g)}\big(E\otimes R^G_{C(g)}(I^G_{C(h)}(F))\big)\\
&\cong I^G_{C(g)}\big(E\otimes I^G_{C(g)}(R^{C(h)}_{C(g)\cap C(h)}(F)\otimes\gamma_{g,h})\big)\\
&\cong I^G_{C(g)}\big(I^{C(g)}_{C(g)\cap C(h)}(R^{C(g)}_{C(g)\cap C(h)}(E)\otimes R^{C(h)}_{C(g)\cap C(h)}(F)\otimes\gamma_{g,h})\big)\\
&\cong I^G_{C(g)\cap C(h)}\big(R^{C(g)}_{C(g)\cap C(h)}(E)\otimes R^{C(h)}_{C(g)\cap C(h)}(F)\otimes\gamma_{g,h}\big)\\
&\cong I^G_{C(gh)}\big(I^{C(gh)}_{C(g)\cap C(h)}(R^{C(g)}_{C(g)\cap C(h)}(E)\otimes R^{C(h)}_{C(g)\cap C(h)}(F)\otimes\gamma_{g,h})\big)\\
&\cong I^G_{C(gh)}(E\star_\omega F).
\end{aligned}$$

Proposition 4.2. *There exists a ring isomorphism*

$$({}^{\omega}K_{\mathrm{orb}}([*/G]),\star_\omega)\cong(R(D^\omega(G)),\otimes).$$

Proof. DPR induction defines a morphism $\phi:\bigoplus_{(g)\subset G}R_\omega(C(g))\to R(D^\omega(G))$. Moreover, we proved above that for $E\in R_{\omega_g}(C(g))$ and $F\in R_{\omega_h}(C(h))$, we have

$$\phi(E)\otimes\phi(F)=\phi(E\star_\omega F);$$

that is, it is a ring homomorphism. By Theorem 4.1, the result follows. □

Corollary 4.3. *The stringy product $\star_\omega$ is associative.*

5. Twisted K-theory for an extraspecial p-group

The goal of this section is to establish a relation between the twisted orbifold K-theories for the orbifolds $[*/H]$ and $[*/G]$, where H is an extraspecial group with exponent p, order p^{2n+1} and $G=(\mathbb{Z}_p)^{2n+1}$. For an odd prime number p, a p-group H is called *extraspecial* if its center $Z(H)$ is a cyclic group of order p, that is $Z(H)\cong\mathbb{Z}_p$, and $H/Z(H)$ is an elementary abelian group. Any extraspecial p-group has order p^{2n+1} for some $n\in\mathbb{N}$. On the other hand, for any n there exist two extraspecial groups of order p^{2n+1} such that a group has exponent p and the other group has exponent p^2. The motivation for these kinds of relations comes from works such as [Goff et al. 2007], where these relations are studied for $p=2$, and to some extent results due to A. Duman [2009]. However, there exists a deeper interest to study these kinds of relations by establishing correspondences with the twisted Drinfeld algebras. In particular, the following result is of utmost importance for obtaining the results of this section:

Theorem 5.1 [Naidu and Nikshych 2008, Corollary 4.20]. *Let H be a finite group $\omega'\in Z^3(H;S^1)$ such that*

- *H contains an abelian normal subgroup K,*
- *$\omega'|_{K\times K\times K}$ is trivial in cohomology (in $H^3(K;S^1)$),*

- *there exists an H-invariant 2-cochain μ over H such that $\delta(\mu)|_{K\times K\times K} = \omega'|_{K\times K\times K}$.*

Then, there exists a group G and an element $\omega \in Z^2(G; S^1)$ such that $R(D^{\omega}(G)) \cong R(D^{\omega'}(H))$.

From the relation established in the previous section between the twisted Drinfeld's algebras and the twisted orbifold K-theory, we get the following corollary under the same assumptions as in the last theorem.

Corollary 5.2. *There exists a ring isomorphism*

$$^{\omega}K_{\mathrm{orb}}([*/G]) \cong {}^{\omega'}K_{\mathrm{orb}}([*/H]).$$

Now, we follow with a nice application of this result.

Proposition 5.3. *Let H be an extraspecial group with order p^{2n+1} and exponent p. Then*

$$K_{\mathrm{orb}}([*/H]) \cong {}^{\omega}K_{\mathrm{orb}}([*/(\mathbb{Z}_p)^{2n+1}])$$

for some nontrivial twisting ω.

Proof. Let H be an extraspecial group. From definition we may assume $K = Z(H) \cong \mathbb{Z}_p$. Now, suppose there exists $\mu \in C^2(H; S^1)$ such that $\delta(\mu)|_{K\times K\times K} = \omega'|_{K\times K\times K}$, which is H-invariant; that is, if we take the action of H on the 2-cochains $C^2(H; S^1)$ defined by ${}^{y}\mu := \mu(yx_1y^{-1}, yx_2y^{-1})$, then ${}^{y}\mu = \mu$ in $C^2(H; S^1)$ for all $y \in H$. Now, since $K = Z(H)$, it follows that ${}^{y}\mu|_K = \mu|_K$ for all $y \in H$. Thus, for all $y \in H$ there exists a 1-chain η_y on H such that $\delta\eta_y = {}^{y}\mu/\mu = 1$. Since K is abelian, we can define the map

$$\nu : H/K \times H/K \to C^1(H; S^1), \quad (y_1, y_1) \mapsto \frac{{}^{y_2}\eta_{y_1}\eta_{y_2}}{\eta_{y_1y_2}}.$$

Lemma 5.4 [Naidu 2007, Lemma 4.2, Corollary 4.3]. *The function ν defines an element in $H^2(H/K; \hat{K}))$.*

However, this element represents a short exact sequence

$$1 \to \hat{K} \to \hat{K} \times_{\nu} H/K \to H/K \to 1,$$

where the product in $\hat{K} \times_{\nu} H/K$ is defined by the formula

$$(\rho_1, x_1)(\rho_2, x_2) := (\nu(x_1, x_2)\rho_1\rho_2, x_1x_2).$$

Now, the element $\omega \in Z^3(G; S^1)$, with $G := \hat{K} \times_{\nu} H/K$, is defined for all (ρ_1, x_1), (ρ_2, x_2), (ρ_3, x_3) in $\hat{K} \times_{\nu} H/K$ by the formula

$$\omega\big((\rho_1, x_1)(\rho_2, x_2)(\rho_3, x_3)\big) := \big(\nu(x_1, x_2)(u(x_3))\big)(1)\rho_1(k_{x_2,x_3}),$$

where $u : H/K \to H$ is a function such that, when composed with the projection

$p : H \to H/K$, yields $p(u(x)) = x$, and $k_{x_2,x_3} \in H$ is an element that satisfies $u(x_1)u(x_2) = k_{x_1,x_2} u(p(u(x_1)u(x_2)))$.

Clearly, when ω' is the trivial 3-cocycle, we can choose μ to be trivial and so ν is also trivial. By definition of an extraspecial group, H/K is an elementary abelian group and if ν is trivial, it follows easily that $\hat{K} \times_\nu H/K \cong (\mathbb{Z}_p)^{2n+1}$. It remains to show that ω is nontrivial in $H^3(G; S^1)$. Take $H = \{h_1, \dots, h_{p^{2n+1}}\}$, $K = Z(H) = \{z_1, \dots, z_p\}$ and $\hat{K} = \{\rho_1, \dots, \rho_p\}$, with ρ_i nontrivial for $i \neq 1$. Denote the quotient group $H/K = \{x_1 K, \dots, x_{p^{2n}} K\}$, with $x_1 = 1_{H/K}$. Now, we define the function $u : H/K \to H$ such that $u(x_i K) = x_i(z_i^{-1})$ for $x_i K \in H/K$, $z_j \in K$. Consider the element $((\rho, x_i K), (\rho, x_i K), (\rho, x_i K))$ with $\rho \in \hat{K}$ fixed and nontrivial. Since ν is trivial, the element ω is reduced to $\rho(k_{x_i K, x_i K}) = \rho(z_i) \neq 1$, which implies that ω is nontrivial. □

Twisted orbifold K-theory for the orbifold $[*/(\mathbb{Z}_p)^n]$**.** With the above result, to calculate the orbifold K-theory structure for $[*/H]$, with H an extraspecial p-group, we only have to calculate the twisted orbifold K-theory for $[*/(\mathbb{Z}_p)^n]$ and a twist element in $H^3((\mathbb{Z}_p)^n; S^1)$, following the constructions presented in Section 3. Because all those constructions are based on the inverse transgression map, we proceed to give an explicit way of calculating it. Later we present an example with a particular twist element, having no trivial inverse transgression map.

*. Inverse transgression map for the group $(\mathbb{Z}_p)^n$ Let us consider the following commutative diagram given by two natural short exact sequences:

(5-1)
$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathbb{Z} & \xrightarrow{\times p} & \mathbb{Z} & \xrightarrow{\pi} & \mathbb{Z}_p & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathbb{Z}_p & \xrightarrow{\times p} & \mathbb{Z}_{p^2} & \xrightarrow{\tau} & \mathbb{Z}_p & \longrightarrow & 0
\end{array}$$

where π and τ are the natural projections, and likewise for the downarrow maps. These two exact sequences in the diagram above induce long exact sequences

(5-2)
$$\cdots \longrightarrow H^{k-1}(BG; \mathbb{Z}_p) \xrightarrow{\partial} H^k(BG; \mathbb{Z}) \xrightarrow{(\times p)_*} H^k(BG; \mathbb{Z}) \xrightarrow{(\pi)_*} H^k(BG; \mathbb{Z}_p) \xrightarrow{\partial} H^{k+1}(BG; \mathbb{Z}) \longrightarrow \cdots,$$

(5-3)
$$\cdots \longrightarrow H^{k-1}(BG; \mathbb{Z}_p) \xrightarrow{\beta} H^k(BG; \mathbb{Z}_p) \xrightarrow{(\times p)_*} H^k(BG; \mathbb{Z}_{p^2}) \xrightarrow{(\tau)_*} H^k(BG; \mathbb{Z}_p) \xrightarrow{\beta} H^{k+1}(BG; \mathbb{Z}_p) \longrightarrow \cdots.$$

Remark 5.5. The connection morphism β of the long exact sequence (5-3) is known as the Bockstein map. It induces a map $\beta : H^*(BG; \mathbb{Z}_p) \to H^*(BG; \mathbb{Z}_p)$

which has the multiplicative property

$$\beta(xy) = \beta(x)y + (-1)^{\deg(x)} x\beta(y).$$

Since G is a p-group, $H^k(BG; -)$ is also a p-group and this implies that the morphism $(\times p)_*$ in the long exact sequences (5-2) and (5-3) is the zero map. Thus, π_* and τ_* are injective maps and $H^k(BG;\mathbb{Z}) \cong H^k(BG;\mathbb{Z}_{p^2})$. On the other hand, by the exactness of the sequence (5-3), we have $H^k(BG;\mathbb{Z}_{p^2}) \cong \operatorname{Ker}\big(\beta : H^k(BG;\mathbb{Z}_p) \to H^{k+1}(BG;\mathbb{Z}_p)\big)$ and then

$$H^k(BG;\mathbb{Z}) \cong \operatorname{Ker}\big(\beta : H^k(BG;\mathbb{Z}_p) \to H^{k+1}(BG;\mathbb{Z}_p)\big).$$

Relation to the inverse transgression map. By definition, the inverse transgression map τ_g is a map defined between the groups $H^k(BG;\mathbb{Z})$ and $H^{k-1}(BC_G(g);\mathbb{Z})$. Since $G = (\mathbb{Z}_p)^n$ is an abelian group, the inverse transgression map can be factorized as

$$\begin{aligned}\tilde{\tau}_g : \operatorname{Ker}\big(\beta : H^k(BG;\mathbb{Z}_p) \to H^{k+1}(BG;\mathbb{Z}_p)\big)&\\ \to \operatorname{Ker}\big(\beta : H^{k-1}(BG;\mathbb{Z}_p) \to H^k(BG;\mathbb{Z}_p)\big)&.\end{aligned}$$

Consider the cohomology ring $H^*(BG;\mathbb{Z}_p) \cong \mathbb{F}_p[x_1,\dots,x_n] \otimes \Lambda[y_1,\dots,y_n]$ with $|x_i| = 2$ and $|y_i| = 1$ for $i = 1,\dots,n$. By the calculations above we need to find a polynomial $p(x_1,\dots,x_n, y_1,\dots,y_n) \in \mathbb{F}_p[x_1,\dots,x_n] \otimes \Lambda[y_1,\dots,y_n]$ of degree k, such that $\beta(p) = 0$ and $\tilde{\tau}_g(p) \neq 0$ for some $g \in G$.

To obtain the desired polynomial, first we do the calculation of the inverse transgression map. Take an element $g = (a_1,\dots,a_n) \in G$ and consider the map

$$G \times \mathbb{Z} \to G \times \langle g \rangle \to G$$

defined by

$$(h, m) \mapsto (h, g^m) \mapsto hg^m.$$

At the level of cohomology we get

$$\begin{aligned}H^*(BG;\mathbb{F}_p) &\to H^*(BG \times B(\mathbb{Z}_p);\mathbb{F}_p) \to H^*(BG \times B\mathbb{Z};\mathbb{F}_p),\\ x_i &\mapsto x_i + a_i w \qquad\qquad \mapsto x_i,\\ y_i &\mapsto y_i + a_i z \qquad\qquad\ \mapsto y_i + a_i z,\end{aligned} \tag{5-4}$$

where

$$\begin{aligned}H^*(BG;\mathbb{F}_p) &= \mathbb{F}_p[x_1,\dots,x_n] \otimes \Lambda[y_1,\dots,y_n],\\ H^*(BG \times B(\mathbb{Z}_p);\mathbb{F}_p) &= \mathbb{F}_p[x_1,\dots,x_n, w] \otimes \Lambda[y_1,\dots,y_n, z],\\ H^*(BG \times B\mathbb{Z};\mathbb{F}_p) &= \mathbb{F}_p[x_1,\dots,x_n] \otimes \Lambda[y_1,\dots,y_n, z].\end{aligned}$$

Now, for the products $x_i y_j, x_i x_j, y_i y_j \in H^*(BG;\mathbb{F}_p)$ we can obtain the calculation of the inverse transgression maps. For the first product $x_i y_j$ we get

$$(x_i y_j) \mapsto (x_i + a_i w)(y_j + a_i z) = x_i y_j + x_i a_j z + a_i w y_j + a_i a_j w z \tag{5-5}$$

in $H^*(BG \times B(\mathbb{Z}_p); \mathbb{F}_p)$ and $(x_i y_j) \mapsto x_i y_j + x_i a_j z$ in $H^*(BG \times B\mathbb{Z}; \mathbb{F}_p)$, from Definition 3.1 it follows that $\tilde{\tau}_g(x_i y_j) = x_i a_j$. For the second product $x_i x_j$ we get

$$(x_i x_j) \mapsto (x_i + a_i w)(x_j + a_i w) = x_i x_j + x_i a_j w + a_i w x_j$$

in $H^*(BG \times B(\mathbb{Z}_p); \mathbb{F}_p)$ and $(x_i x_j) \mapsto x_i x_j$ in $H^*(BG \times B\mathbb{Z}; \mathbb{F}_p)$; hence

$$\tilde{\tau}_g(x_i x_j) = 0. \tag{5-6}$$

Finally, for the product $y_i y_j$ we get

$$(y_i y_j) \mapsto (y_i + a_i z)(y_j + a_i z) = y_i y_j + y_i a_j z + a_i z y_j \tag{5-7}$$

in $H^*(BG \times B(\mathbb{Z}_p); \mathbb{F}_p)$ and $(y_i y_j) \mapsto y_i y_j + (a_j y_i - a_i y_j) z$ in $H^*(BG \times B\mathbb{Z}; \mathbb{F}_p)$; therefore $\tilde{\tau}_g(y_i y_j) = (a_j y_i - a_i y_j)$.

Since we are interested in calculating the inverse transgression map for elements $\alpha \in H^4(G; \mathbb{Z})$, we consider only polynomials of degree 4 in

$$H^*(BG; \mathbb{F}_p) = \mathbb{F}_p[x_1, \dots, x_n] \otimes \Lambda[y_1, \dots, y_n].$$

Now, we present some examples of the inverse transgression map. It is easiest to consider the cases $n = 2$ and $n = 3$. In the first case the inverse transgression map is a trivial map. In the latter the inverse transgression map is more interesting.

Example 5.6. $\underline{n = 2}$. For $p \neq 2$ we have $H^*(BG; \mathbb{F}_p) = \mathbb{F}_p[x_1, x_2] \otimes \Lambda[y_1, y_2]$ with $|y_i| = 1$ and $|\beta y_i| = |x_i| = 2$. Thus, we can just consider linear combinations of the polynomials $p_1(x_1, x_2, y_1, y_2) := x_1 x_2$, $p_2(x_1, x_2, y_1, y_2) := x_1 y_1 y_2$ and $p_3(x_1, x_2, y_1, y_2) := x_2 y_1 y_2$. For p_1 the calculations leading up to (5-6) show that $\tilde{\tau}_g(p_1) = 0$. Thus we need to find a $(\mathbb{Z}_p)$-linear combination p of the polynomials p_2 and p_3 such that $\beta(p) = 0$. But we have

$$\begin{aligned}\beta(p_3) &= x_2(\beta(y_1) y_2 - y_1 \beta(y_2)) = x_2(x_1 y_2 - y_1 x_2),\\ \beta(p_2) &= x_1(\beta(y_1) y_2 - y_1 \beta(y_2)) = x_1(x_1 y_2 - y_1 x_2).\end{aligned}$$

Therefore, there does not exists such a $(\mathbb{Z}_p)$-linear combination.

$\underline{n = 3}$. By analyzing the degree of the polynomials, we obtain the element

$$p(x_1, x_2, x_3, y_1, y_2, y_3) = x_1 y_2 y_3 - x_2 y_1 y_3 + x_3 y_1 y_2, \tag{5-8}$$

which satisfies the condition $\beta(p) = 0$. To check this, we use the property of β

noted in Remark 5.5:

$$\begin{aligned}
\beta(p) &= \beta(x_1y_2y_3)-\beta(x_2y_1y_3)+\beta(x_3y_1y_2)\\
&= \beta(x_1)y_2y_3+x_1\beta(y_2y_3)-\beta(x_2)y_1y_3-x_2\beta(y_1y_3)+\beta(x_3)y_1y_2+x_3\beta(y_1y_2)\\
&= x_1\beta(y_2)y_3-x_1y_2\beta(y_3)-x_2\beta(y_1)y_3+x_2y_1\beta(y_3)+x_3\beta(y_1)y_2-x_3y_1\beta(y_2)\\
&= x_1x_2y_3-x_1y_2x_3-x_2x_1y_3+x_2y_1x_3+x_3x_1y_2-x_3y_1x_2\\
&= 0.
\end{aligned}$$

The inverse transgression map for an element $g=(a_1,a_2,a_3)\in(\mathbb{Z}_p)^3$ evaluated in the polynomial p gives

$$\begin{aligned}
\tau_g(p) &= \tau_g(x_1y_2y_3)-\tau_g(x_2y_1y_3)+\tau_g(x_3y_1y_2)\\
&= x_1(a_3y_2-a_2y_3)-x_2(a_3y_1-a_1y_3)+x_3(a_2y_1-a_1y_2)\\
&= a_1(x_2y_3-x_3y_2)+a_2(x_3y_1-x_1y_3)+a_3(x_1y_2-x_2y_1).
\end{aligned} \tag{5-9}$$

Lemma 5.7. *Let $g=(a_1,a_2,a_3)$ and $h=(b_1,b_2,b_3)$ be elements in $G=(\mathbb{Z}_p)^3$. The double inverse transgression map of p is equal to*

$$\tau_h\tau_g(p)=[(a_1,a_2,a_3)\times(b_1,b_2,b_3)]\cdot(x_1,x_2,x_3). \tag{5-10}$$

Remark 5.8. With $n=3$, this example shows that for $n\geq 3$ the inverse transgression map is nontrivial. We can always consider the element $p(x_1,\dots,x_n,y_1,\dots,y_n)=x_iy_jy_k-x_jy_iy_k+x_ky_iy_j$ as being in $H^4((Z_p)^n;\mathbb{Z})$. By calculations similar to those leading to (5-9), we can prove that $\beta(p)=0$ while $\tau_g(p)\neq 0$ for $g\in(\mathbb{Z}_p)^n$.

By using the inverse transgression map for the group $(\mathbb{Z}_p)^n$ presented above and by the decomposition formula presented in Theorem 3.6 in [Becerra and Uribe 2009], we can calculate the explicit structure of the twisted orbifold K-theory for the orbifold $[*/(\mathbb{Z}_p)^3]$ and the twist element α as the element in $H^3((\mathbb{Z}_p)^3;S^1)$ associated to the polynomial defined in (5-8) via the isomorphism $H^3((\mathbb{Z}_p)^3;S^1)\cong H^4((\mathbb{Z}_p)^3;\mathbb{Z})$. Note that in this case

$${}^{\alpha}K_{\mathrm{orb}}([*/(\mathbb{Z}_p)^3])=\bigoplus_{g\in(\mathbb{Z}_p)^3}{}^{\alpha_g}K_{(\mathbb{Z}_p)^3}(*)\cong\bigoplus_{g\in(\mathbb{Z}_p)^3}R_{\alpha_g}((\mathbb{Z}_p)^3).$$

Now, for each $g\in\mathbb{Z}_p$, the decomposition formula implies

$$R_{\alpha_g}((\mathbb{Z}_p)^3)\otimes\mathbb{Q}\cong\prod_{g,h\in(\mathbb{Z}_p)^3}\left(\mathbb{Q}(\zeta_p)_{h,\alpha_g}\right)^{(\mathbb{Z}_p)^3},$$

where ζ_p is a p-root of the unity. Note that the action of $(\mathbb{Z}_p)^3$ on $\mathbb{Q}(\zeta_p)_{h,\alpha_g}$ is to multiply by the double inverse transgression map evaluated on $k\in(\mathbb{Z}_p)^3$, $\tau_h(\alpha_g)(k)$.

By Lemma 5.7, we get

$$\big(\mathbb{Q}(\zeta_p)_{h,\alpha_g}\big)^{(\mathbb{Z}_p)^3} = \begin{cases} \mathbb{Q}(\zeta_p) & \text{if } g = \lambda h,\ \lambda \in \mathbb{Z}_p, \\ 0 & \text{else.} \end{cases} \tag{5-11}$$

So, for $h \neq 0$, we have

$$R_{\alpha_g}((\mathbb{Z}_p)^3) \otimes \mathbb{Q} = \prod_{\lambda \in \mathbb{Z}_p} \mathbb{Q}(\zeta_p),$$

while for $g = 0$, we get

$$R_{\alpha_1}((\mathbb{Z}_p)^3) \otimes \mathbb{Q} = \prod_{\lambda \in (\mathbb{Z}_p)^3} \mathbb{Q}(\zeta_p).$$

Then, the twisted orbifold K-theory module for the orbifold $[*/(\mathbb{Z}_p)^3]$ turns out to be

$${}^{\alpha}K_{\mathrm{orb}}([*/(\mathbb{Z}_p)^3]) \otimes \mathbb{Q} = \prod_{\lambda \in \mathbb{Z}_p} \mathbb{Q}(\zeta_p) \ \oplus \prod_{\lambda \in (\mathbb{Z}_p)^3} \mathbb{Q}(\zeta_p)$$

and the product structure is defined via the product of the elements in $\mathbb{Q}(\zeta_p)$.

6. Final remarks

With the result presented in Section 4 about the Grothendieck ring associated to the semigroup of representations of the twisted Drinfeld double $D^{\omega}(G)$ and the twisted orbifold K-theory, we found a nice relation between two structures coming from different sources. As we already said, the orbifold $[*/G]$ is a particular case of a more general kind of orbifolds obtained by the almost free action of a compact Lie group G on a compact manifold M. With a little more structure, the stringy product introduced in Section 3 can be extended to a stringy product on the module ${}^{\alpha}K_{\mathrm{orb}}([M/G])$ (in the same way as in [Becerra and Uribe 2009]), where $[M/G]$ denotes the orbifold structure obtained by the almost free action (see [Adem and Ruan 2003] for the details of this structure). Therefore, under suitable hypotheses we can think about the twisted orbifold K-theory ${}^{\omega}K_{\mathrm{orb}}([M/G])$ as a more general object which coincides with the Grothendieck ring $R(D^{\omega}(G))$ if G is a finite group and $M = \{*\}$. Nevertheless, we shall explore the interpretation and consequences of this more general object. Next, we focus our attention on the results obtained in Section 5, where we establish an explicit relation between the twisted orbifold K-theories of the orbifolds $[*/H]$ and $[*/(\mathbb{Z}_p)^n]$, where H is a particular extraspecial p-group. In the same spirit, we look for some general relation between the twisted orbifold K-theories ${}^{\alpha}K_{\mathrm{orb}}([M/G])$ and ${}^{\beta}K_{\mathrm{orb}}([M/K])$ of the orbifolds $[M/G]$ and $[M/K]$, for suitable twistings $\alpha \in H^3(G; S^1)$ and $\beta \in H^3(H; S^1)$, and appropriate actions of the finite groups G and K on a compact manifold M. In the same way, we hope that some analogous results may be obtained if G and K are compact Lie groups acting almost freely on a compact manifold M. By our preliminary observations, in order to obtain such results, some hypothesis on the almost free actions of the compact Lie groups G and K must be added.

References

[Adem and Ruan 2003] A. Adem and Y. Ruan, "Twisted orbifold K-theory", *Comm. Math. Phys.* **237**:3 (2003), 533–556. MR 2004e:19004 Zbl 1051.57022

[Adem et al. 2007] A. Adem, Y. Ruan, and B. Zhang, "A stringy product on twisted orbifold K-theory", *Morfismos* **11**:2 (2007), 33–64. arXiv math/0605534

[Becerra and Uribe 2009] E. Becerra and B. Uribe, "Stringy product on twisted orbifold K-theory for abelian quotients", *Trans. Amer. Math. Soc.* **361**:11 (2009), 5781–5803. MR 2011b:19008 Zbl 1191.14068

[Carey and Wang 2008] A. L. Carey and B.-L. Wang, "Thom isomorphism and push-forward map in twisted K-theory", *J. K-Theory* **1**:2 (2008), 357–393. MR 2009g:55005 Zbl 1204.55006

[tom Dieck 1987] T. tom Dieck, *Transformation groups*, de Gruyter Studies in Mathematics **8**, de Gruyter, Berlin, 1987. MR 89c:57048 Zbl 0611.57002

[Dijkgraaf et al. 1991] R. Dijkgraaf, V. Pasquier, and P. Roche, "Quasi Hopf algebras, group cohomology and orbifold models", *Nuclear Phys. B Proc. Suppl.* **18**:2 (1991), 60–72. MR 92m:81238 Zbl 0957.81670

[Drinfel'd 1987] V. G. Drinfel'd, "Quantum groups", pp. 798–820 in *Proceedings of the International Congress of Mathematicians* (Berkeley, CA, 1986), vol. 1, edited by A. M. Gleason, Amer. Math. Soc., Providence, RI, 1987. MR 89f:17017 Zbl 0667.16003

[Goff et al. 2007] C. Goff, G. Mason, and S.-H. Ng, "On the gauge equivalence of twisted quantum doubles of elementary abelian and extra-special 2-groups", *J. Algebra* **312**:2 (2007), 849–875. MR 2008d:16057 Zbl 1171.16021

[Karpilovsky 1993] G. Karpilovsky, *Group representations*, vol. 2, North-Holland Mathematics Studies **177**, North-Holland, Amsterdam, 1993. MR 94f:20001 Zbl 0778.20001

[Nabi Duman 2009] A. Nabi Duman, "The fusion algebra of an extraspecial p-group", *J. Math. Phys.* **50**:8 (2009), Article ID #082301. MR 2010h:20023 Zbl 1203.81077

[Naidu 2007] D. Naidu, "Categorical Morita equivalence for group-theoretical categories", *Comm. Algebra* **35**:11 (2007), 3544–3565. MR 2008j:18007 Zbl 1143.18009

[Naidu and Nikshych 2008] D. Naidu and D. Nikshych, "Lagrangian subcategories and braided tensor equivalences of twisted quantum doubles of finite groups", *Comm. Math. Phys.* **279**:3 (2008), 845–872. MR 2009b:16092 Zbl 1139.16028

[Segal 1968] G. Segal, "Equivariant K-theory", *Inst. Hautes Études Sci. Publ. Math.* **34** (1968), 129–151. MR 38 #2769 Zbl 0199.26202

[Vafa 2001] C. Vafa, "Superstrings and topological strings at large N", *J. Math. Phys.* **42**:7 (2001), 2798–2817. MR 2002h:81227 Zbl 1060.81594

[Wasserman 1969] A. G. Wasserman, "Equivariant differential topology", *Topology* **8** (1969), 127–150. MR 40 #3563 Zbl 0215.24702

[Willerton 2008] S. Willerton, "The twisted Drinfeld double of a finite group via gerbes and finite groupoids", *Algebr. Geom. Topol.* **8**:3 (2008), 1419–1457. MR 2009g:57050 Zbl 1154.57029

[Witherspoon 1996] S. J. Witherspoon, "The representation ring of the twisted quantum double of a finite group", *Canad. J. Math.* **48**:6 (1996), 1324–1338. MR 97m:16078 Zbl 0876.16027

[Witten 1998] E. Witten, "D-branes and K-theory", *J. High Energy Phys.* **12** (1998), Article ID #19. MR 2000e:81151 Zbl 0959.81070

Received May 16, 2012. Revised November 28, 2012.

MARIO VELÁSQUEZ
UNIVERSIDAD DE LOS ANDES
DEPARTAMENTO DE MATEMÁTICAS
CARRERA 1 NO 18A–12 EDIFICIO H
BOGOTÁ
COLOMBIA
ma.velasquez109@uniandes.edu.co

EDWARD BECERRA
DEPARTMENT OF MATHEMATICS
UNIVERSIDAD NACIONAL DE COLOMBIA
CARRERA 30 CALLE 45, CIUDAD UNIVERSITARIA
BOGOTÁ
COLOMBIA
esbecerrar@unal.edu.co

HERMES MARTINEZ
ESCUELA DE MATEMÁTICAS
UNIVERSIDAD SERGIO ARBOLEDA
CALLE 74 NO. 14–14
BOGOTÁ
COLOMBIA
hermes.martinez@ima.usergioarboleda.edu.co

PACIFIC JOURNAL OF MATHEMATICS
Vol. 264, No. 2, 2013
dx.doi.org/10.2140/pjm.2013.264.491

LINEAR RESTRICTION ESTIMATES FOR THE WAVE EQUATION WITH AN INVERSE SQUARE POTENTIAL

JUNYONG ZHANG AND JIQIANG ZHENG

We study modified linear restriction estimates associated with the wave equation with an inverse square potential. In particular, we show that the classical linear restriction estimates hold in their almost sharp range when the initial data is radial.

1. Introduction and statement of main result

In this paper, we study a modified restriction estimate associated with the wave equation perturbed by an inverse square potential. More precisely, we consider the following wave equation with a singular potential:

$$\begin{cases} \partial_t^2 u - \Delta u + \dfrac{a}{|x|^2} u = 0, & (t,x) \in \mathbb{R} \times \mathbb{R}^n,\ a \in \mathbb{R}, \\ u(t,x)|_{t=0} = 0, & \partial_t u(t,x)|_{t=0} = f(x). \end{cases} \tag{1-1}$$

The scale-covariant elliptic operator $P_a := -\Delta + a/|x|^2$ appearing in (1-1) plays a key role in many problems of physics and geometry. The heat and Schrödinger flows for the elliptic operator P_a have been studied in the theory of combustion [Vazquez and Zuazua 2000] and in quantum mechanics [Kalf et al. 1975]. The wave equation (1-1) arises in the study of the wave propagation on conic manifolds [Cheeger and Taylor 1982]. There has been a lot of interest in developing Strichartz estimates both for the Schrödinger and wave equations with the inverse square potential; we refer the reader to [Burq et al. 2003; 2004; Planchon et al. 2003b; 2003a; Miao et al. 2013b]. However, as far as we know, there are few results about the restriction estimates associated with the operator P_a arising in the study of eigenfunction estimates of P_a. Here, we address some restriction issues in special settings associated with the operator P_a.

In the case $a = 0$ — the linear wave equation with no potential — we can solve

Partially supported by the Fundamental Research Foundation of B.I.T (20111742015).
MSC2010: 42B37, 35Q40, 47J35.
Keywords: linear restriction estimate, spherical harmonics, Bessel function, inverse square potential.

the equation by the Fourier transform formula

$$u(t,x)=\frac{\sin(t\sqrt{-\Delta})}{\sqrt{-\Delta}}f=\frac{1}{2i}\int_{\mathbb{R}^n}e^{2\pi i x\cdot\xi}\big(e^{2\pi i t|\xi|}-e^{-2\pi i t|\xi|}\big)\hat{f}(\xi)\frac{d\xi}{|\xi|},\tag{1-2}$$

where the Fourier transform is defined by

$$\hat{f}(\xi)=\int_{\mathbb{R}^n}e^{-2\pi i x\cdot\xi}f(x)\,dx.$$

It is well known that the spacetime norm estimate of $u(t,x)$ is connected with the linear adjoint cone restriction estimate

$$\|(F\,d\sigma)^\vee\|_{L^q_{t,x}(\mathbb{R}\times\mathbb{R}^n)}\le C_{p,q,n,S}\|F\|_{L^p(S,d\sigma)},\tag{1-3}$$

where F is a Schwartz function and the inverse spacetime Fourier transform of the measure $F\,d\sigma$ is defined by

$$(F\,d\sigma)^\vee(t,x)=\int_S F(\tau,\xi)e^{2\pi i(x\cdot\xi+t\tau)}\,d\sigma(\xi)=\int_{\mathbb{R}^n}F(|\xi|,\xi)e^{2\pi i(x\cdot\xi+t|\xi|)}\frac{d\xi}{|\xi|}.$$

Here, the set S is a nonempty smooth compact subset of the cone

$$\{(\tau,\xi)\in\mathbb{R}\times\mathbb{R}^n:\tau=|\xi|\}\quad\text{with}\quad n\ge 2.$$

The canonical measure $d\sigma$ is the pull back of the measure $d\xi/|\xi|$ under the projection map $(\tau,\xi)\mapsto\xi$. By the decay of $(d\sigma)^\vee$ and the Knapp counterexample, the two necessary conditions for (1-2) are

$$q>\frac{2n}{n-1}\quad\text{and}\quad\frac{n+1}{q}\le\frac{n-1}{p'}\tag{1-4}$$

(see [Stein 1979; Tao 2003a]). The corresponding linear adjoint restriction conjecture for cones asserts that:

Conjecture 1.1. *The inequality* (1-3) *holds with constants depending on n, p, q, and S if and only if the inequalities* (1-4) *are satisfied.*

Even though there is a large amount of literature focused on this problem, it remains open for $n\ge 4$. For progress on this conjecture, we refer the readers to [Taberner 1985; Strichartz 1977; Tao 2001; 2003a; 2003b; Tao et al. 1998; Wolff 2001]. Shao [2009a] provided two simple and novel arguments to prove that Conjecture 1.1 holds true for the spatial rotation invariant functions which are supported on the cone. Motivated by [Shao 2009a], Miao et al. [2012] utilized expansions in spherical harmonics and analyzed the asymptotic behavior of the Bessel function to generalize Shao's result for cone cases by establishing, assuming (1-4), that

$$\|(F\,d\sigma)^\vee\|_{L^q_t(\mathbb{R};L^q_{r^{n-1}\,dr}L^2_\theta(\mathbb{S}^{n-1}))}\le C_{p,q,n,S}\|F\|_{L^p(S,d\sigma)}.$$

In the case $a \neq 0$, the spacetime Fourier transform is no longer so useful; one can instead establish an approximate parametrix for the fundamental solution and try to obtain good control over it. In our case we resort to expansions in spherical harmonics and Hankel transforms, for technical reasons involving the singular potential; compare [Burq et al. 2003; Planchon et al. 2003b; Miao et al. 2013b]. Although the harmonic expansion expression leads to some loss of angular regularity in the restriction estimates, it allows us to show the restriction estimates when q is close to $2n/(n-1)$. A key ingredient in this process is to explore the oscillatory properties of the Bessel function and $e^{it|\xi|}$ to overcome the difficulties arising from the low decay of the Bessel function $J_\nu(r)$ when $\nu \sim r$. Finally, by using the properties of the hypergeometric function shown in [Planchon et al. 2003b], we prove an inequality involving the Hankel transform to obtain the desired result.

Main Theorem. *Assume $n \geq 2$ and $a > -\frac{1}{4}(n-2)^2$, and let u be the solution of* (1-1). *Suppose that $p > 1$ and*

$$q = \frac{p'(n+1)}{n-1} > \frac{2n}{n-1}. \tag{1-5}$$

Then there exists a constant C, depending only on p, q, n, and a, satisfying the following conditions:

(i) *If f is a radial Schwartz function, then*

$$\|u(t,x)\|_{L^q_{t,x}(\mathbb{R}\times\mathbb{R}^n)} \leq C \left\| |\xi|^{-\frac{1}{p}} \hat{f} \right\|_{L^p(\mathbb{R}^n)}. \tag{1-6}$$

(ii) *If f is a Schwartz function* (*may not be radial*) *and $p \geq 2$, then*

$$\|u(t,x)\|_{L^q_t(\mathbb{R};L^q_{r^{n-1}\,dr}L^2_\theta(\mathbb{S}^{n-1}))} \leq C \left\| |\xi|^{-\frac{1}{p}} \hat{f} \right\|_{L^p(\mathbb{R}^n)}. \tag{1-7}$$

Remarks. (i) This extends the classical restriction estimate associated with the Laplace operator to a restriction estimate associated with $-\Delta + a/|x|^2$. We obtain more estimates than the Strichartz estimates of [Burq et al. 2003; Planchon et al. 2003b], which focus on $p = 2$. The theorem can also be viewed as an extension of the result in [Chen et al. 2012] about the operator $-\Delta + a/r^2$ acting on $L^2\big((0,\infty); r^{n-1}\,dr\big)$.

(ii) The theorem means that we almost show that the classical linear restriction estimates hold for radial functions in the conjecture range.

(iii) When $a = 0$, we recover the cone restriction result in [Shao 2009b]. When supp $\hat{f}$ is compact, we can extend the result to $q \geq p'(n+1)/(n-1)$, which is the same range as in the cone restriction conjecture.

(iv) Equation (1-6) gives a Strichartz-type estimate

$$\|u(t,x)\|_{L^{2(n+1)/(n-1)}_{t,x}(\mathbb{R}\times\mathbb{R}^n)} \leq C \left\| |\nabla|^{-\frac{1}{2}} f \right\|_{L^2(\mathbb{R}^n)}$$

for the radial solution. The method used here generalizes the result for the radial initial data to a linear finite combination of products of the Hankel transform of radial functions and spherical harmonics. We hope to remove the whole radial assumption in (1-6) in the future, at least for $q \geq 2(n+3)/(n+1)$.

(v) If $\hat{f} \subset \{\xi : N \leq |\xi| \leq 2N\}$ and f is radial, the method here can be employed to obtain the Strichartz estimate

$$\|u(t,x)\|_{L^q_{t,x}(\mathbb{R}\times\mathbb{R}^n)} \leq CN^{\frac{n-2}{2}-\frac{n+1}{q}}\|f\|_{L^2(\mathbb{R}^n)} \quad \text{for } q > \frac{2n}{n-1}.$$

We remark that the pair (q,q) is allowed to be out of the admissible requirement in [Planchon et al. 2003b], it is however consistent with the admissible range due to [Miao et al. 2013b].

(vi) We rely heavily on the harmonic expansion formula to give the expression of the solution due to the potential, which causes the restriction $p \geq 2$. It is possible that the resolvent expression can be used to remove this restriction.

Now we introduce some notation. We use $A \lesssim B$ to denote the statement that $A \leq CB$ for some constant C, which may vary from line to line and depend on various parameters. We write $A \sim B$ to mean that $A \lesssim B \lesssim A$.

If the constant C depends on parameters other than p, q, n, and S, we denote this fact explicitly using subscripts. For instance, C_ϵ should be understood as a positive constant depending on ϵ in addition to (possibly) p, q, n, and S.

Pairs of conjugate indices are written as p and p', where

$$\frac{1}{p}+\frac{1}{p'}=1 \quad \text{and} \quad 1 \leq p \leq \infty.$$

This paper is organized as follows: In Section 2, we present some simple facts about the Hankel transforms and the Bessel functions and also recall the van der Corput lemma. Section 3 is devoted to the proof of the Main Theorem via expansions in spherical harmonics and an analysis of the asymptotic behavior of the Bessel function. Finally, in an Appendix, we show an inequality used in Section 3, involving the Hankel transforms.

2. Preliminaries

Before turning to Hankel transforms and Bessel functions, we recall the expansion formula in spherical harmonics. For details, refer to [Stein and Weiss 1971]. For convenience, we write

$$\xi = \rho\omega \quad \text{and} \quad x = r\theta \quad \text{with } \omega, \theta \in \mathbb{S}^{n-1}.$$

We denote by $\mathcal{H}^k$ the space of spherical harmonics of degree k on $\mathbb{S}^{n-1}$, whose dimension is given by

$$d(0)=1, \qquad d(k)=\frac{2k+n-2}{k}C_{n+k-3}^{k-1}\simeq\langle k\rangle^{n-2} \quad \text{for } k>0.$$

Note that if $n=2$ this dimension is 2 for all k.

Any $g\in L^2(\mathbb{R}^n)$ can be expanded in spherical harmonics as

$$g(x)=\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}a_{k,l}(r)Y_{k,l}(\theta), \tag{2-1}$$

where $\{Y_{k,1},\dots,Y_{k,d(k)}\}$ is an orthogonal basis of $\mathscr{H}^k$. We have the orthogonal decomposition

$$L^2(\mathbb{S}^{n-1})=\bigoplus_{k=0}^{\infty}\mathscr{H}^k,$$

and by orthogonality,

$$\|g(x)\|_{L^2_\theta}=\|a_{k,l}(r)\|_{l^2_{k,l}}. \tag{2-2}$$

The Hankel transform formula (see Theorem 3.10 in [Stein and Weiss 1971], for instance) relates the Fourier transform of g to spherical harmonics. It reads

$$\hat{g}(\rho\omega)=\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}2\pi i^kY_{k,l}(\omega)\rho^{-\frac{n-2}{2}}\int_0^\infty J_{k+\frac{n-2}{2}}(2\pi r\rho)a_{k,l}(r)r^{\frac{n}{2}}\,dr. \tag{2-3}$$

Here the Bessel function $J_k(r)$ of order k is defined by

$$J_k(r)=\frac{\left(\frac{r}{2}\right)^k}{\Gamma\left(k+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}\int_{-1}^{1}e^{isr}\left(1-s^2\right)^{\frac{2k-1}{2}}\,ds, \quad \text{with } k>-\tfrac{1}{2} \text{ and } r>0.$$

A simple computation gives the rough estimate

$$|J_k(r)|\le\frac{Cr^k}{2^k\Gamma\left(k+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}\left(1+\frac{1}{k+\frac{1}{2}}\right), \tag{2-4}$$

where C is an absolute constant. These estimates will be mainly used when $r\lesssim 1$. Another well known asymptotic expansion about the Bessel function is

$$J_k(r)=r^{-\frac{1}{2}}\sqrt{\frac{2}{\pi}}\cos\left(r-\frac{k\pi}{2}-\frac{\pi}{4}\right)+O_k(r^{-3/2}) \quad \text{as } r\to\infty, \tag{2-5}$$

but with a constant depending on k (see [Stein and Weiss 1971]). As pointed out in [Stein 1993], if one seeks a uniform bound for large r and k, then the best one can do is $|J_k(r)|\le Cr^{-\frac{1}{3}}$. To investigate the asymptotic behavior in k and r, we recall

Schläfli's integral representation [Watson 1944] of the Bessel function. For $r \in \mathbb{R}^+$ and $k > -\frac{1}{2}$,

$$J_k(r) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ir\sin\theta - ik\theta}\,d\theta - \frac{\sin(k\pi)}{\pi}\int_0^\infty e^{-(r\sinh s + ks)}\,ds \tag{2-6}$$
$$=: \tilde{J}_k(r) - E_k(r).$$

We remark that $E_k(r) = 0$ when $k \in \mathbb{Z}^+$. A simple computation gives, for $r > 0$,

$$|E_k(r)| = \left|\frac{\sin(k\pi)}{\pi}\int_0^\infty e^{-(r\sinh s + ks)}\,ds\right| \le C(r+k)^{-1}. \tag{2-7}$$

Next, we recall some properties of the Bessel function $J_k(r)$ from [Stein 1993; Stempak 2000]; see [Miao et al. 2013a] for a detailed proof.

Lemma 2.1 (asymptotics of the Bessel function). *Assume $k \gg 1$. Let $J_k(r)$ be the Bessel function of order k defined as above. Then there exist a large constant C and a small constant c independent of k and r such that*

$$|J_k(r)| \le Ce^{-c(k+r)} \qquad \textit{when } r \le k/2, \tag{2-8}$$

$$|J_k(r)| \le Ck^{-\frac{1}{3}}\big(k^{-\frac{1}{3}}|r-k|+1\big)^{-\frac{1}{4}} \qquad \textit{when } k/2 \le r \le 2k, \tag{2-9}$$

$$J_k(r) = r^{-\frac{1}{2}}\sum_{\pm} a_\pm(r)e^{\pm ir} + E(r) \qquad \textit{when } r \ge 2k, \tag{2-10}$$

where $|a_\pm(r)| \le C$ and $|E(r)| \le Cr^{-1}$.

We define

$$\mu(k) = \frac{n-2}{2} + k, \quad \nu(k) = \sqrt{\mu^2(k) + a} \quad \text{with } a > -\frac{(n-2)^2}{4}. \tag{2-11}$$

For the sake of simplicity, we sometimes write ν instead of $\nu(k)$. Let g be a Schwartz function defined on $\mathbb{R}^n$. We define the Hankel transform of order ν:

$$(\mathscr{H}_\nu g)(\rho\omega) = \int_0^\infty (r\rho)^{-\frac{n-2}{2}} J_\nu(r\rho) g(r\omega) r^{n-1}\,dr. \tag{2-12}$$

If the function g is radial, we can drop the dependence on ω from both sides.

We remark that if g has the expansion (2-1), it follows from (2-3) that

$$\hat{g}(\xi) = \sum_{k=0}^\infty \sum_{l=1}^{d(k)} 2\pi i^k Y_{k,l}(\omega)(\mathscr{H}_{\mu(k)} a_{k,l})(\rho). \tag{2-13}$$

The following properties of the Hankel transform are proved in [Burq et al. 2003; Planchon et al. 2003b]:

Lemma 2.2. *Let $\mathcal{H}_\nu$ be as above and let*

$$A_{\nu(k)} := -\partial_r^2 - \frac{n-1}{r}\partial_r + \Big(\nu^2(k) - \Big(\frac{n-2}{2}\Big)^2\Big) r^{-2}.$$

(i) $\mathcal{H}_\nu = \mathcal{H}_\nu^{-1}$.

(ii) $\mathcal{H}_\nu$ *is self-adjoint*: $\mathcal{H}_\nu = \mathcal{H}_\nu^*$.

(iii) $\mathcal{H}_\nu$ *is an L^2 isometry*: $\|\mathcal{H}_\nu \phi\|_{L^2_\xi} = \|\phi\|_{L^2_x}$.

(iv) $\mathcal{H}_\nu(A_\nu \phi)(\xi) = |\xi|^2 (\mathcal{H}_\nu \phi)(\xi)$, *for* $\phi \in L^2$.

We conclude this section by recalling van der Corput's lemma [Stein 1993]:

Lemma 2.3. *Let ϕ be a smooth real-valued function defined on an interval $[a, b]$, and assume $|\phi^{(k)}(x)| \geq 1$ for all $x \in [a, b]$. Assume moreover that either $k \geq 2$, or $k = 1$ and $\phi'(x)$ is monotonic. Then*

$$\left| \int_a^b e^{i\lambda\phi(x)}\, dx \right| \leq c_k \lambda^{-\frac{1}{k}}, \tag{2-14}$$

with c_k independent of ϕ and λ.

3. Proof of the main theorem

In this section, we will use the asymptotic properties of the Bessel function and the stationary phase argument to establish two estimates for the Hankel transform. A key ingredient is to effectively exploit the oscillatory property of the Bessel function and $e^{it|\xi|}$ to obtain more decay.

The Hankel transform and the solution. Let us consider (1-1) in polar coordinates. Write $v(t, r, \theta) = u(t, r\theta)$ and $g(r, \theta) = f(r\theta) = f(x)$. Then $v(t, r, \theta)$ satisfies

$$\begin{cases} \partial_{tt} v - \partial_{rr} v - \dfrac{n-1}{r}\partial_r v - \dfrac{1}{r^2}\Delta_\theta v + \dfrac{a}{r^2} v = 0, \\ v(0, r, \theta) = 0, \quad \partial_t v(0, r, \theta) = g(r, \theta). \end{cases} \tag{3-1}$$

We use the spherical harmonic expansion to write

$$g(r, \theta) = \sum_{k=0}^{\infty} \sum_{l=1}^{d(k)} a_{k,l}(r) Y_{k,l}(\theta). \tag{3-2}$$

Using separation of variables, we can write v as a superposition

$$v(t, r, \theta) = \sum_{k=0}^{\infty} \sum_{l=1}^{d(k)} v_{k,l}(t, r) Y_{k,l}(\theta), \tag{3-3}$$

where $v_{k,l}$ satisfies

$$\begin{cases} \partial_{tt} v_{k,l} - \partial_{rr} v_{k,l} - \dfrac{n-1}{r}\partial_r v_{k,l} + \dfrac{k(k+n-2)+a}{r^2} v_{k,l} = 0, \\ v_{k,l}(0,r) = 0, \quad \partial_t v_{k,l}(0,r) = a_{k,l}(r) \end{cases} \tag{3-4}$$

for each $k, l \in \mathbb{N}$, $1 \le l \le d(k)$. Define

$$A_{\nu(k)} := -\partial_r^2 - \frac{n-1}{r}\partial_r + \frac{\nu^2(k) - \left(\frac{n-2}{2}\right)^2}{r^2}. \tag{3-5}$$

Then we are reduced to considering the system

$$\begin{cases} \partial_{tt} v_{k,l} + A_{\nu(k)} v_{k,l} = 0, \\ v_{k,l}(0,r) = 0, \quad \partial_t v_{k,l}(0,r) = a_{k,l}(r). \end{cases} \tag{3-6}$$

Applying the Hankel transform to (3-6), we have, by Lemma 2.2,

$$\begin{cases} \partial_{tt} \tilde{v}_{k,l} + \rho^2 \tilde{v}_{k,l} = 0, \\ \tilde{v}_{k,l}(0,\xi) = 0, \quad \partial_t \tilde{v}_{k,l}(0,\xi) = b_{k,l}(\rho), \end{cases} \tag{3-7}$$

where

$$\tilde{v}_{k,l}(t,\rho) = (\mathcal{H}_\nu v_{k,l})(t,\rho), \quad b_{k,l}(\rho) = (\mathcal{H}_\nu a_{k,l})(\rho). \tag{3-8}$$

Solving this ODE and using the Hankel transform, we obtain

$$\begin{aligned} v_{k,l}(t,r) &= \int_0^\infty (r\rho)^{-\frac{n-2}{2}} J_{\nu(k)}(r\rho) \tilde{v}_{k,l}(t,\rho) \rho^{n-2}\, d\rho \\ &= \frac{1}{2i} \int_0^\infty (r\rho)^{-\frac{n-2}{2}} J_{\nu(k)}(r\rho) \big(e^{it\rho} - e^{-it\rho}\big) b_{k,l}(\rho) \rho^{n-2}\, d\rho. \end{aligned}$$

Therefore, we get

$$\begin{aligned} u(x,t) &= v(t,r,\theta) \\ &= \sum_{k=0}^\infty \sum_{l=1}^{d(k)} Y_{k,l}(\theta) \int_0^\infty (r\rho)^{-\frac{n-2}{2}} J_{\nu(k)}(r\rho) \sin(t\rho) b_{k,l}(\rho) \rho^{n-2}\, d\rho \\ &= \sum_{k=0}^\infty \sum_{l=1}^{d(k)} Y_{k,l}(\theta) \mathcal{H}_{\nu(k)}\big[\rho^{-1} \sin(t\rho) b_{k,l}(\rho)\big](r). \end{aligned} \tag{3-9}$$

Estimates of Hankel transforms. We now turn to some key estimates needed for proving the main theorem.

Proposition 3.1. *Let $R \gg 1$ and let φ be a smooth function supported in the interval $I := \left[\frac{1}{2}, 1\right]$ and taking values in* $[0,1]$. *Then*

$$\left\| \left(\sum_{k=0}^{\infty} \sum_{l=1}^{d(k)} \left| \int_0^{\infty} e^{-it\rho} J_{\nu(k)}(r\rho) b_{k,l}(\rho) \varphi(\rho)\, d\rho \right|^2 \right)^{\frac{1}{2}} \right\|_{L_t^2(\mathbb{R}; L_r^2([R,2R]))} \leq C \left\| \left(\sum_{k=0}^{\infty} \sum_{l=1}^{d(k)} |b_{k,l}(\rho)|^2 \right)^{\frac{1}{2}} \right\|_{L_\rho^2(I)}, \tag{3-10}$$

where C is a constant independent of R.

Proof. Using the Plancherel theorem in t, we have

$$\text{LHS of (3-10)} \lesssim \left\| \left(\sum_{k=0}^{\infty} \sum_{l=1}^{d(k)} \| J_{\nu(k)}(r\rho) b_{k,l}(\rho) \varphi(\rho) \|_{L_\rho^2}^2 \right)^{\frac{1}{2}} \right\|_{L_r^2([R,2R])}. \tag{3-11}$$

With this, it is easy to verify (3-10) if we can prove that

$$\int_R^{2R} |J_k(r)|^2\, dr \leq C, \tag{3-12}$$

where $R \gg 1$ and C is independent of k and R. To prove (3-12), we write

$$\int_R^{2R} |J_k(r)|^2\, dr = \int_{I_1} |J_k(r)|^2\, dr + \int_{I_2} |J_k(r)|^2\, dr + \int_{I_3} |J_k(r)|^2\, dr, \tag{3-13}$$

where

$$\begin{aligned} I_1 &= [R, 2R] \cap [0, k/2], \\ I_2 &= [R, 2R] \cap [k/2, 2k], \\ I_3 &= [R, 2R] \cap [2k, \infty]. \end{aligned}$$

By using (2-8) and (2-10) in Lemma 2.1, we have

$$\int_{I_1} |J_k(r)|^2\, dr \leq C \int_{I_1} e^{-cr} r\, dr \leq C e^{-cR}, \quad \int_{I_3} |J_k(r)|^2\, dr \leq C. \tag{3-14}$$

For the remaining interval, we write

$$\int_{I_2} |J_k(r)|^2\, dr \leq \int_{\frac{k}{2}}^{2k} |J_k(r)|^2\, dr \leq C \int_{\frac{k}{2}}^{2k} k^{-\frac{2}{3}} \big(1 + k^{-\frac{1}{3}} |r - k|\big)^{-\frac{1}{2}}\, dr \leq C,$$

where the last inequality follows from the fact that the integral is uniformly bounded (by $2 + \sqrt{2}$) for all $k > 0$. Together with (3-14), this yields (3-12). $\square$

Proposition 3.2. *Suppose $R \gg 1$. Let φ be a smooth function supported in the interval $I := \left[\frac{1}{2}, 1\right]$ and taking values in* $[0, 1]$.

(i) *If K is finite, there exists a constant C_K independent of R such that*

$$\left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} \left| \int_0^\infty e^{-it\rho} J_{\nu(k)}(r\rho) b_{k,l}(\rho) \varphi(\rho)\, d\rho \right|^2 \right)^{\frac{1}{2}} \right\|_{L_t^\infty(\mathbb{R}; L_r^\infty([R,2R]))} \leq C_K R^{-\frac{1}{2}} \left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} |b_{k,l}(\rho)|^2 \right)^{\frac{1}{2}} \right\|_{L^1_\rho(I)}. \tag{3-15}$$

(ii) *If K is infinite, there exists a constant C independent of R such that*

$$\left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} \left| \int_0^\infty e^{-it\rho} J_{\nu(k)}(r\rho) b_{k,l}(\rho) \varphi(\rho)\, d\rho \right|^2 \right)^{\frac{1}{2}} \right\|_{L_t^\infty(\mathbb{R}; L_r^\infty([R,2R]))} \leq C R^{-\frac{1}{2}} \left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} |b_{k,l}(\rho)|^2 \right)^{\frac{1}{2}} \right\|_{L^2_\rho(I)}. \tag{3-16}$$

Proof. We first prove (3-15). Recalling (2-5) we can write $|J_{\nu(k)}(r)| \leq C_K r^{-\frac{1}{2}}$ when $r \gg 1$. By the Minkowski inequality and the Hausdorff–Young inequality in t, there exists a constant C_K independent of R such that

$$\left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} \left| \int_0^\infty e^{-it\rho} J_{\nu(k)}(r\rho) b_{k,l}(\rho) \varphi(\rho)\, d\rho \right|^2 \right)^{\frac{1}{2}} \right\|_{L_t^\infty(\mathbb{R}; L_r^\infty([R,2R]))} \leq C_K R^{-\frac{1}{2}} \left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} \left|b_{k,l}(\rho)\right|^2 \right)^{\frac{1}{2}} \right\|_{L^1_\rho(I)}.$$

This proves (3-15). When K is infinite, we need to show a precise estimate uniform in K. We utilize the Schläfli's integral representation of the Bessel function (2-6) to write $J_{\nu(k)}(r\rho) = E_{\nu(k)}(r\rho) + \tilde{J}_{\nu(k)}(r\rho)$. By (2-7), the Minkowski inequality, and the Hausdorff–Young inequality in t, there exists a constant C independent of K and R such that

$$\left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} \left| \int_0^\infty e^{-it\rho} E_{\nu(k)}(r\rho) b_{k,l}(\rho) \varphi(\rho)\, d\rho \right|^2 \right)^{\frac{1}{2}} \right\|_{L_t^\infty(\mathbb{R}; L_r^\infty([R,2R]))} \leq C R^{-1} \left\| \left(\sum_{k=0}^{K} \sum_{l=1}^{d(k)} \left|b_{k,l}(\rho)\right|^2 \right)^{\frac{1}{2}} \right\|_{L^1_\rho(I)}.$$

Thus, it remains to prove (3-16) with $J_{\nu(k)}$ replaced by $\tilde{J}_{\nu(k)}$. We consider $0<\delta \ll 1$ to be fixed later, and write $[-\pi, \pi] = I_1 \cup I_2 \cup I_3$, with

$$I_1 = \{\theta : |\theta| \leq \delta\},$$
$$I_2 = [-\pi, -\pi/2-\delta] \cup [\pi/2+\delta, \pi],$$
$$I_3 = [-\pi, \pi] \setminus (I_1 \cup I_2).$$

We define

$$\Phi_{r,k}(\theta)=\sin\theta-\frac{k\theta}{r}, \tag{3-17}$$

and let $\chi_\delta(\theta)$ be a smooth function satisfying

$$\chi_\delta(\theta)=\begin{cases}1, & \theta\in[-\delta,\delta],\\ 0, & \theta\notin[-2\delta,2\delta].\end{cases} \tag{3-18}$$

Then write

$$\tilde{J}_k(r)=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{ir\Phi_{r,k}(\theta)}\,d\theta=\tilde{J}_k^1(r)+\tilde{J}_k^2(r)+\tilde{J}_k^3(r), \tag{3-19}$$

with

$$\begin{aligned}\tilde{J}_k^1(r)&:=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{ir\Phi_{r,k}(\theta)}\chi_\delta(\theta)\,d\theta,\\ \tilde{J}_k^2(r)&:=\frac{1}{2\pi}\int_{I_2}e^{ir\Phi_{r,k}(\theta)}\,d\theta,\\ \tilde{J}_k^3(r)&:=\frac{1}{2\pi}\int_{I_3}e^{ir\Phi_{r,k}(\theta)}(1-\chi_\delta(\theta))\,d\theta.\end{aligned}$$

When $\theta\in I_2$, the function $\Phi'_{r,k}(\theta)=\cos\theta-k/r$ is monotonic in the intervals $[-\pi,-\pi/2-\delta]$ and $[\pi/2+\delta,\pi]$, and satisfies

$$|\Phi'_{r,k}(\theta)|\geq\frac{k}{r}+|\cos\theta|\geq\sin\delta.$$

Then van der Corput's lemma (Lemma 2.3) gives, uniformly in k,

$$\left|\frac{1}{2\pi}\int_{I_2}e^{ir\Phi_{r,k}(\theta)}\,d\theta\right|\leq c_\delta r^{-1}. \tag{3-20}$$

When $\theta\in I_3$, we have $|\Phi''_{r,k}(\theta)|\geq\sin\delta$, and Lemma 2.3 again yields that

$$\left|\frac{1}{2\pi}\int_{I_3}e^{ir\Phi_{r,k}(\theta)}(1-\chi_\delta(\theta))\,d\theta\right|\leq c_\delta r^{-\frac{1}{2}} \tag{3-21}$$

uniformly in k.

Using arguments similar to those above, it follows from (3-20) and (3-21) that

$$\begin{aligned}\bigg\|\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\bigg|\int_0^\infty e^{-it\rho}\big(\tilde{J}^2_{\nu(k)}(r\rho)+\tilde{J}^3_{\nu(k)}(r\rho)\big)b_{k,l}(\rho)\varphi(\rho)\,d\rho\bigg|^2\bigg)^{\frac{1}{2}}\bigg\|_{L_t^\infty(\mathbb{R};L_r^\infty([R,2R]))}&\\ \lesssim R^{-\frac{1}{2}}\bigg\|\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\big|b_{k,l}(\rho)\big|^2\bigg)^{\frac{1}{2}}\bigg\|_{L^1_\rho(I)}&.\end{aligned}$$

To establish (3-16) with $J_{\nu(k)}$ replaced by $\tilde{J}^1_{\nu(k)}$, we need to use the oscillation of

$e^{it\rho}$ effectively. To this end, we write the Fourier series as $b_{k,l}(\rho)=\sum_j b_{k,l}^j e^{i\frac{\pi}{2}\rho j}$, where

$$b_{k,l}^j=\frac{1}{4}\int_0^4 e^{-i\frac{\pi}{2}\rho j}b_{k,l}(\rho)\,d\rho. \tag{3-22}$$

Then $\sum_j |b_{k,l}^j|^2=\|b_{k,l}(\rho)\|^2_{L^2_\rho(I)}$. For simplicity, we use the scaling argument to reduce the problem by replacing t and r by $2\pi t$ and $2\pi r$ respectively, and define

(3-23)
$$\psi^k_{t-j/4}(r)=\frac{1}{2\pi}\int_0^\infty e^{-2\pi i(t-\frac{j}{4})\rho}\int_{\mathbb{R}} e^{2\pi i\rho r\sin\theta-i\nu(k)\theta}\chi_\delta(\theta)\,d\theta\varphi(\rho)\,d\rho.$$

Let $m=t-j/4$. Then we write

$$\begin{aligned}\psi^k_m(r)&=\frac{1}{2\pi}\int_{\mathbb{R}^2} e^{2\pi i\rho(r\sin\theta-m)}e^{-i\nu(k)\theta}\chi_\delta(\theta)\varphi(\rho)\,d\rho\,d\theta\\&=\frac{1}{2\pi}\int_{\mathbb{R}}\check{\varphi}(r\sin\theta-m)e^{-i\nu(k)\theta}\chi_\delta(\theta)\,d\theta.\end{aligned} \tag{3-24}$$

For our purpose, we need to investigate the asymptotic behavior of the function $\psi^k_m(r)$. We consider two subcases:

(a) $4R\le |m|$. $R\ge 1$, hence $|m|\ge 4$. Since $\check{\varphi}$ is a Schwartz function, we have

$$|\check{\varphi}(r\sin\theta-m)|\le C_N(1+|r\sin\theta-m|)^{-N}\quad\text{for all } N>0.$$

On the other hand, we have

$$|r\sin\theta-m|\ge |m|-r|\sin\theta|\ge \tfrac{1}{100}|m|,$$

since $r\le 2R\le |m|$ and $|\theta|\le 2\delta$. Thus, (3-24) gives

$$|\psi^k_m(r)|\le C_{\delta,N}(1+|m|)^{-N}. \tag{3-25}$$

Keeping in mind that $m=t-j/4$, we have

$$\begin{aligned}&\Bigg\|\bigg(\sum_{k=0}^\infty\sum_{l=1}^{d(k)}\bigg|\sum_{\substack{j:\\ 4R\le|t-j/4|}} b^j_{k,l}\psi^k_{t-j/4}(r)\bigg|^2\bigg)^{\frac12}\Bigg\|_{L^\infty_t(\mathbb{R};L^\infty_r([R,2R]))}\\&\le C_{\delta,N}R^{-N}\Bigg\|\bigg(\sum_{k=0}^\infty\sum_{l=1}^{d(k)}\bigg|\sum_{\substack{j:\\ 4R\le|t-j/4|}}|b^j_{k,l}|(1+|t-\tfrac{j}{4}|)^{-N}\bigg|^2\bigg)^{\frac12}\Bigg\|_{L^\infty_t(\mathbb{R};L^\infty_r([R,2R]))}.\end{aligned}$$

By the Cauchy–Schwarz inequality, and choosing N large enough, the above is bounded by

$$C_{\delta,N}R^{-N}\bigg\|\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\sum_{j}|b_{k,l}^{j}|^2(1+|t-\tfrac{j}{4}|)^{-N}\bigg)^{\frac12}\bigg\|_{L_t^\infty(\mathbb{R};L_r^\infty([R,2R]))}$$
$$\leq C_{\delta,N}R^{-N}\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\sum_{j}|b_{k,l}^{j}|^2\bigg)^{\frac12}\lesssim R^{-N}\bigg\|\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\big|b_{k,l}(\rho)\big|^2\bigg)^{\frac12}\bigg\|_{L^2_\rho(I)}.$$

(b) $|m|<4R$. Again, since $\check{\varphi}$ is a Schwartz function,

$$|\check{\varphi}(r\sin\theta-m)|\leq C_N(1+|r\sin\theta-m|)^{-N}\quad\text{for all }\ N>0. \tag{3-26}$$

By (3-24), this gives

$$\big|\psi_m^k(r)\big|\leq\frac{C_N}{2\pi}\bigg(\int_{\substack{\{\theta:|\theta|<2\delta,\\ |r\sin\theta-m|\geq1\}}}d\theta+\int_{\substack{\{\theta:|\theta|<2\delta,\\ |r\sin\theta-m|\geq1\}}}(1+|r\sin\theta-m|)^{-N}\,d\theta\bigg).$$

Let $y=r\sin\theta-m$; then

$$\big|\psi_m^k(r)\big|\leq\frac{C_N}{2\pi r}\bigg(\int_{\{y:|y|\leq1\}}dy+\int_{\{y:|y|\geq1\}}(1+|y|)^{-N}\,dy\bigg)\lesssim\frac{1}{r}. \tag{3-27}$$

For fixed t, R, we define the set $A=\{j\in\mathbb{Z}:|t-j/4|\leq4R\}$. It is easy to see the cardinality of A is $O(R)$. Thus, it follows from (3-27) and the Cauchy–Schwarz inequality that

$$\bigg\|\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\bigg|\sum_{j\in A}b_{k,l}^{j}\psi_{t-j/4}^{k}(r)\bigg|^2\bigg)^{\frac12}\bigg\|_{L_t^\infty(\mathbb{R};L_r^\infty([R,2R]))}$$
$$\leq C_{\delta,N}R^{-\frac12}\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\sum_{j}|b_{k,l}^{j}|^2\bigg)^{\frac12}\lesssim R^{-\frac12}\bigg\|\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\big|b_{k,l}(\rho)\big|^2\bigg)^{\frac12}\bigg\|_{L^2_\rho(I)}\qquad\square$$

Proposition 3.3. *Let φ be a smooth function supported on $I=\left[\frac12,1\right]$ and taking values in* $[0,1]$, *and let R be a positive real number. Assume* (1-5) *is satisfied*, *and consider the quantity*

$$Q=\bigg\|r^{-\frac{n-2}{2}}\bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\bigg|\int_0^\infty e^{-it\rho}J_{\nu(k)}(r\rho)b_{k,l}(\rho)\varphi(\rho)\,d\rho\bigg|^2\bigg)^{\frac12}\bigg\|_{L_t^q(\mathbb{R};L^q_{r^{n-1}dr}([R,2R]))}.$$

(*Recall the definition of $\nu=\nu(k)$ in* (2-11).)

(i) *When K is finite*, *there exists a constant C_K independent of R such that*

$$Q\leq C_K\min\big\{R^{\frac{n}{q}},R^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\big\}\bigg\|\bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}|b_{k,l}(\rho)|^2\bigg)^{\frac12}\bigg\|_{L^p_\rho(I)}. \tag{3-28}$$

(ii) *When K is infinite*, *there exists a constant C independent of R such that*

$$Q \le C \min\left\{R^{\frac{n}{q}}, R^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\right\}\left\|\left(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}|b_{k,l}(\rho)|^2\right)^{\frac{1}{2}}\right\|_{L^2_\rho(I)}. \tag{3-29}$$

Proof. We first consider the case $R \lesssim 1$. The Minkowski inequality and the Hausdorff–Young inequality in t show that

$$Q \lesssim \left\| r^{-\frac{n-2}{2}}\left(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\left\|J_{\nu(k)}(r\rho)b_{k,l}(\rho)\varphi(\rho)\right\|^2_{L^{q'}_\rho}\right)^{\frac{1}{2}}\right\|_{L^q_{r^{n-1}\,dr}([R,2R])}.$$

Hence, by (2-4), there exists a constant C independent of K such that

$$Q \le C\left(\int_R^{2R} r^{-\frac{(n-2)q}{2}}\left(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\left|\frac{(8\pi r)^\nu}{2^\nu\Gamma\left(\nu+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)}\right|^2\left\|\varphi(\rho)b_{k,l}(\rho)\right\|^2_{L^{q'}_\rho}\right)^{\frac{q}{2}} r^{n-1}\,dr\right)^{\frac{1}{q}}$$

and

$$Q \le CR^{\frac{n}{q}}\left\|\left(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\left|\varphi(\rho)b_{k,l}(\rho)\right|^2\right)^{\frac{1}{2}}\right\|_{L^{q'}_\rho(I)} r.$$

Secondly, we consider the case $R \gg 1$. By Prepositions 3.1 and 3.2, we use interpolation to obtain

$$\left\| r^{-\frac{n-2}{2}}\left(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\left|\int_0^\infty J_\nu(r\rho)e^{it\rho}b_{k,l}(\rho)\varphi(\rho)\,d\rho\right|^2\right)^{\frac{1}{2}}\right\|_{L^q_t(\mathbb{R};L^q_{r^{n-1}\,dr}([R,2R]))}$$
$$\le C_K R^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\left\|\left(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\left|\varphi(\rho)b_{k,l}(\rho)\right|^2\right)^{\frac{1}{2}}\right\|_{L^{q'}_\rho(I)}.$$

When K is infinite,

$$\left\| r^{-\frac{n-2}{2}}\left(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\left|\int_0^\infty J_\nu(r\rho)e^{it\rho}b_{k,l}(\rho)\varphi(\rho)\,d\rho\right|^2\right)^{\frac{1}{2}}\right\|_{L^q_t(\mathbb{R};L^q_{r^{n-1}\,dr}([R,2R]))}$$
$$\le CR^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\left\|\left(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\left|\varphi(\rho)b_{k,l}(\rho)\right|^2\right)^{\frac{1}{2}}\right\|_{L^2_\rho(I)}.$$

In view of (1-5) and since $\operatorname{supp}\varphi \subset \left[\frac{1}{2}, 1\right]$, this shows (3-28) and (3-29). □

Conclusion of the proof of the Main Theorem. We know that if f is radial, so is u in (3-9). To prove the Main Theorem, we need to estimate the following by (3-9):

$$\left\|\sum_{k=0}^{K}\sum_{l=1}^{d(k)} Y_{k,l}(\theta)\mathcal{H}_\nu\left(\rho^{-1}\sin(t\rho)b_{k,l}(\rho)\right)(r)\right\|_{L^q_t(\mathbb{R};L^q_{r^{n-1}\,dr}L^2_\theta(\mathbb{S}^{n-1}))} \tag{3-30}$$

in the cases of $K = 0$ and $K = \infty$, which correspond to the radial case and the general case respectively. To this end, we use orthogonality and apply a dyadic decomposition to (3-30) to obtain the estimate

$$\begin{aligned}
&(3\text{-}30)\\
&\leq C\Bigg(\sum_R\bigg(\sum_M\bigg\|\bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)} J_\nu(r\rho)e^{it\rho}\chi\Big(\frac{\rho}{M}\Big)\\
&\qquad\qquad\times\bigg|\int_0^\infty(r\rho)^{-\frac{n-2}{2}}b_{k,l}(\rho)\rho^{n-2}\,d\rho\bigg|^2\bigg)^{\frac12}\bigg\|_{L_t^q(\mathbb{R};L^q_{r^{n-1}dr}([R,2R]))}\bigg)^q\Bigg)^{\frac1q},\\
&=:\Sigma
\end{aligned}$$

where R and M are dyadic numbers and χ is a smooth function supported on $\left[\frac12,1\right]$ and taking values in $[0,1]$. By a scaling argument, we have

$$\begin{aligned}
\Sigma \leq C\Bigg(\sum_R\bigg(\sum_M M^{(n-1)-\frac{n+1}{q}}\bigg\|\bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\\
\bigg|\int_0^\infty(r\rho)^{-\frac{n-2}{2}}J_\nu(r\rho)e^{it\rho}\chi(\rho)b_{k,l}(M\rho)\rho^{n-2}d\rho\bigg|^2\bigg)^{\frac12}\bigg\|_{L_t^q(\mathbb{R};L^q_{r^{n-1}dr}([RM,2RM]))}\bigg)^q\Bigg)^{\frac1q}.
\end{aligned}$$

Applying Proposition 3.3 with $\varphi(\rho) = \chi(\rho)\rho^{\frac{n}{2}-1}$ to the above, one can see that when K is finite,

$$\begin{aligned}
\Sigma \leq C_K\Bigg(\sum_R\bigg(\sum_M \min\Big\{(RM)^{\frac{n}{q}},(RM)^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\Big\}\\
\times M^{(n-1)-\frac{n+1}{q}}\bigg\|\bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\big|\chi(\rho)\rho^{\frac{n}{2}-1}b_{k,l}(M\rho)\big|^2\bigg)^{\frac12}\bigg\|_{L^p_\rho}\bigg)^q\Bigg)^{\frac1q},
\end{aligned}$$

and when K is infinite,

$$\begin{aligned}
\Sigma \leq C\Bigg(\sum_R\bigg(\sum_M \min\Big\{(RM)^{\frac{n}{q}},(RM)^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\Big\}\\
\times M^{(n-1)-\frac{n+1}{q}}\bigg\|\bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\big|\chi(\rho)\rho^{\frac{n}{2}-1}b_{k,l}(M\rho)\big|^2\bigg)^{\frac12}\bigg\|_{L^2_\rho}\bigg)^q\Bigg)^{\frac1q}.
\end{aligned}$$

Since $q > 2n/(n-1)$, one has

$$\sup_R\sum_M \min\Big\{(RM)^{\frac{n}{q}},(RM)^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\Big\} < \infty,$$

$$\sup_M\sum_R \min\Big\{(RM)^{\frac{n}{q}},(RM)^{-\frac{n-1}{2}\left(1-\frac{2n}{q(n-1)}\right)}\Big\} < \infty.$$

Then by Schur's test lemma and the embedding $l^p \hookrightarrow l^q$ with $q > \dfrac{2n}{n-1} > p$, we have, in the case when K is finite,

$$\begin{aligned}\Sigma &\le C_K\Bigg(\sum_M M^{\left((n-1)-\frac{n+1}{q}\right)p}\Bigg\|\chi(\rho)\Bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}|b_{k,l}(M\rho)|^2\Bigg)^{\frac12}\Bigg\|^p_{L^p_\rho}\Bigg)^{\frac1p}\\ &\le C_K\Bigg(\sum_M M^{\left(\frac{n-1}{p'}-\frac{n+1}{q}\right)p}\Bigg\|\chi\Big(\frac{\rho}{M}\Big)\Bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\big|b_{k,l}(\rho)\big|^2\Bigg)^{\frac12}\rho^{\frac{n-2}{p}}\Bigg\|^p_{L^p_\rho}\Bigg)^{\frac1p}\\ &\le C_K\Bigg\|\Bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\big|b_{k,l}(\rho)\big|^2\Bigg)^{\frac12}\Bigg\|_{L^p_{\rho^{n-2}\,d\rho}(\mathbb{R}^+)},\end{aligned}$$

and in the case when K is infinite and $p \ge 2$,

$$\begin{aligned}\Sigma &\le C\Bigg(\sum_M M^{\left(\frac{n-1}{p'}-\frac{n+1}{q}\right)q}\Bigg\|\chi\Big(\frac{\rho}{M}\Big)\Bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\big|b_{k,l}(\rho)\big|^2\Bigg)^{\frac12}\rho^{\frac{n-2}{2}}\Bigg\|^q_{L^p_\rho}\Bigg)^{\frac1q}\\ &\le C\Bigg\|\Bigg(\sum_{k=0}^{\infty}\sum_{l=1}^{d(k)}\big|b_{k,l}(\rho)\big|^2\Bigg)^{\frac12}\Bigg\|_{L^p_{\rho^{n-2}\,d\rho}(\mathbb{R}^+)}.\end{aligned}$$

By Lemma 2.2, we have $b_{k,l}(\rho) = \mathcal{H}_{\nu(k)}\mathcal{H}_{\mu(k)}[\mathcal{H}_{\mu(k)}a_{k,l}](\rho)$. To proceed we make use of the following fact, whose proof we defer to the Appendix:

Claim. *For the measure space* $(\mathbb{R}^+, dw(\rho))$, *where* $dw(\rho) = \rho^{n-2}\,d\rho$, *and for* $1 < p < \infty$, *denote by* $L^p(w)$ *the corresponding Lebesgue space equipped with the norm*

$$\|f\|_p = \left(\int_0^\infty |f|^p\,dw\right)^{\frac1p}.$$

Let $\mathcal{H}_{\nu(k)}, \mathcal{H}_{\mu(k)}$ *be the Hankel transforms defined above and suppose that*

$$\frac{n-2}{2} - \nu(0) < \frac{n-1}{p} < \frac{n-2}{2} + \mu(0) + 2. \tag{3-31}$$

Then there exists a constant C such that, for any $\{f_k\}_{k=0}^{\infty} \in L^p(w; l^2)$, *we have*

$$\Bigg\|\Bigg(\sum_k |\mathcal{H}_{\nu(k)}\mathcal{H}_{\mu(k)} f_k|^2\Bigg)^{\frac12}\Bigg\|_{L^p(w)} \le C\Bigg\|\Bigg(\sum_k |f_k|^2\Bigg)^{\frac12}\Bigg\|_{L^p(w)}. \tag{3-32}$$

Condition (3-31) is satisfied because $1 < p < \dfrac{2n}{n-1}$ by the Main Theorem's assumptions. Thus, applying the Claim we get

$$\Sigma \le C\Bigg\|\Bigg(\sum_{k=0}^{K}\sum_{l=1}^{d(k)}\big|[\mathcal{H}_{\mu(k)}a_{k,l}](\rho)\big|^2\Bigg)^{\frac12}\Bigg\|_{L^p_{\rho^{n-2}\,d\rho}(\mathbb{R}^+)}.$$

By (2-2) and (2-13), under the conditions of the Main Theorem, we further have

$$\Sigma \le C\, \big\| |\xi|^{-\frac{1}{p}} \hat{f}(\xi) \big\|_{L^p_{\rho^{n-1}\,d\rho}(\mathbb{R}^+; L^2_\omega(\mathbb{S}^{n-1}))} \le C\, \big\| |\xi|^{-\frac{1}{p}} \hat{f}(\xi) \big\|_{L^p_\xi(\mathbb{R}^n)}.$$

This completes the proof. □

Appendix: Proof of (3-32)

Let $T_k = \mathcal{H}_{\mu(k)}\mathcal{H}_{\nu(k)}$, and set $\lambda = \dfrac{n-2}{2}$. We first show that

$$\|T_k f_k\|_{L^p(w)} \le C \|f_k\|_{L^p(w)}, \tag{A-1}$$

by following the argument used to prove Theorem 3.1 in [Planchon et al. 2003b]. By that argument, we can write

$$(T_k f_k)(\rho) = \int_0^\infty k^0_{\nu,\mu(k)}(\rho, s) f_k(s) s^{n-1}\, ds,$$

where the kernel is given by

$$k^0_{\alpha,\beta}(\rho, s) = \begin{cases} A_{\alpha,\beta} \dfrac{s^{\beta-\lambda}}{\rho^{\lambda+\beta+2}} F\Big(\dfrac{\alpha+\beta}{2} + 1, \dfrac{\beta-\alpha}{2} + 1; \beta + 1; \Big(\dfrac{s}{\rho}\Big)^2\Big) & \text{for } s < \rho, \\ A_{\beta,\alpha} \dfrac{s^{\alpha-\lambda}}{\rho^{\lambda+\alpha+2}} F\Big(\dfrac{\beta+\alpha}{2} + 1, \dfrac{\alpha-\beta}{2} + 1; \alpha + 1; \Big(\dfrac{\rho}{s}\Big)^2\Big) & \text{for } s > \rho, \end{cases}$$

where $F(a, b; c; d)$ is the hypergeometric function and

$$A_{\alpha,\beta} = \frac{2\Gamma(\frac{\alpha+\beta}{2} + 1)}{\Gamma(\frac{\beta-\alpha}{2})\Gamma(\beta + 1)}.$$

When s is near ρ, the kernel $k^0_{\nu,\mu}(\rho, s)$ behaves like $c(\rho-s)^{-1} + O(-\log|\rho-s|)$. Define

$$(\widetilde{T}_k[s^{\frac{n-1}{p}} f_k(s)])(\rho) := \int_0^\infty \widetilde{k}^0_{\nu,\mu}(\rho, s)\big[s^{\frac{n-1}{p}} f_k(s)\big]\frac{ds}{s},$$

where

$$\widetilde{k}^0_{\nu,\mu}(\rho, s) = \rho^{\frac{n-1}{p}} k^0_{\nu,\mu}(\rho, s) s^{n-\frac{n-1}{p}}.$$

Then

$$\big(\widetilde{T}_k[s^{\frac{n-1}{p}} f_k(s)]\big)(\rho) = \int_0^\infty \rho^{\frac{n-1}{p}} k^0_{\nu,\mu}(\rho, s) s^{n-\frac{n-1}{p}} [s^{\frac{n-1}{p}} f_k(s)]\frac{ds}{s} = \rho^{\frac{n-1}{p}} (T_k f_k)(\rho).$$

Note that

$$\big\|(T_k f_k)(\rho)\big\|_{L^p(w)} = \big\|\rho^{\frac{n-1}{p}} (T_k f_k)(\rho)\big\|_{L^p_{\rho^{-1}\,d\rho}}.$$

To prove (A-1), it suffices to show

$$\|\widetilde{T}_k f_k\|_{L^p(\rho^{-1}\,d\rho)} \le C \|f_k\|_{L^p(\rho^{-1}\,d\rho)}. \tag{A-2}$$

Again by the argument in [Planchon et al. 2003b], one has

$$|k^0_{\nu,\mu}(\rho,s)| = \begin{cases} O(\rho^{-\lambda-\mu-2+\epsilon}s^{-\lambda+\mu-\epsilon}) & \text{for } s<\rho,\\ O(\rho^{\nu-\lambda-\epsilon}s^{-\lambda-\nu-2+\epsilon}) & \text{for } s>\rho.\end{cases}$$

Then

$$|\widetilde{k}^0_{\nu,\mu}(\rho,s)| = \begin{cases} O\big((s/\rho)^{\lambda+\mu+2-\frac{n-1}{p}-\epsilon}\big) & \text{for } s<\rho,\\ O\big((s/\rho)^{\lambda-\nu-\frac{n-1}{p}+\epsilon}\big) & \text{for } s>\rho.\end{cases}$$

Since $\lambda-\nu<(n-1)/p<\lambda+\mu+2$, the kernel $\widetilde{k}^0_{\nu,\mu}$ is bounded in $L^1(d\rho/\rho)$. Using logarithmic coordinates, we express the operator T_k as a convolution operator with the kernel $\widetilde{k}^0_{\nu,\mu}$. When $s\sim\rho$, we recall that $\widetilde{k}^0_{\nu,\mu}$ is a Calderón–Zygmund kernel behaving like

$$c(\rho-s)^{-1}+O(-\log|\rho-s|).$$

Applying Young's inequality to the region away from $\rho\sim s$ and Calderón–Zygmund theory to the region $\rho\sim s$, we obtain (A-2), and so (A-1).

By a similar argument, we show the adjoint operator T_k^* is also bounded in $L^{p'}(w)$, provided $\lambda-\mu<(n-1)/p'<\lambda+\nu+2$, which is true for $1<p<\infty$.

Now we are ready to show (3-32). We consider two cases:

If $1<p\le 2$, since the adjoint operator $T_k^*=\mathcal{H}_{\nu(k)}\mathcal{H}_{\mu(k)}$ is bounded in $L^{p'}(w)$, we get (3-32) by duality.

If $2\le p<\infty$, we set $q:=p/2$ (forgetting the earlier value of q). Then $q\ge 1$, and we have

$$\begin{aligned}\Big\|\Big(\sum_k |T_kf_k|^2\Big)^{\frac12}\Big\|^2_{L^p(w)} &= \Big\|\sum_k |T_kf_k|^2\Big\|_{L^q(w)}\\ &= \sup_{\substack{g\ge 0\\ g\in L^{q'}(w)}} \Big|\int_0^\infty \sum_k |T_kf_k|^2 g(\rho)\rho^{n-2}\,d\rho\Big|\\ &= \sum_k \sup_{\substack{g\ge 0\\ g\in L^{q'}(w)}} \int_0^\infty |T_kf_k|^2 g(\rho)\rho^{n-2}\,d\rho\\ &= \sum_k \|T_kf_k\|^2_{L^p(w)}.\end{aligned}$$

By (A-1), we see that

$$\begin{aligned}\Big\|\Big(\sum_k |T_kf_k|^2\Big)^{\frac12}\Big\|^2_{L^p(w)} &\le C\sum_k \|f_k\|^2_{L^p(w)}\\ &= C\sum_k \sup_{\substack{g\ge 0\\ g\in L^{q'}(w)}} \Big|\int_0^\infty |f_k|^2 g(\rho)\rho^{n-2}\,d\rho\Big|\end{aligned}$$

$$\le C \sup_{\substack{g\ge 0\\ g\in L^{q'}(w)}} \int_0^\infty \sum_k |f_k|^2 g(\rho)\rho^{n-2}\,d\rho$$

$$\le C\Big\|\sum_k |f_k|^2\Big\|^2_{L^q(w)} \le C\Big\|\Big(\sum_k |f_k|^2\Big)^{\frac{1}{2}}\Big\|^2_{L^p(w)}.$$

References

[Burq et al. 2003] N. Burq, F. Planchon, J. G. Stalker, and A. S. Tahvildar-Zadeh, "Strichartz estimates for the wave and Schrödinger equations with the inverse-square potential", *J. Funct. Anal.* **203**:2 (2003), 519–549. MR 2004m:35025 Zbl 1030.35024

[Burq et al. 2004] N. Burq, F. Planchon, J. G. Stalker, and A. S. Tahvildar-Zadeh, "Strichartz estimates for the wave and Schrödinger equations with potentials of critical decay", *Indiana Univ. Math. J.* **53**:6 (2004), 1665–1680. MR 2005k:35241 Zbl 1084.35014

[Cheeger and Taylor 1982] J. Cheeger and M. Taylor, "On the diffraction of waves by conical singularities, I", *Comm. Pure Appl. Math.* **35**:3 (1982), 275–331. MR 84h:35091a Zbl 0526.58049

[Chen et al. 2012] P. Chen, E. M. Ouhabaz, A. Sikora, and L. Yan, "Restriction estimates, sharp spectral multipliers and endpoint estimates for Bochner–Riesz means", preprint, 2012. arXiv 1202.4052v1

[Kalf et al. 1975] H. Kalf, U.-W. Schmincke, J. Walter, and R. Wüst, "On the spectral theory of Schrödinger and Dirac operators with strongly singular potentials", pp. 182–226 in *Spectral theory and differential equations* (Dundee, 1974), edited by W. N. Everitt, Lecture Notes in Mathematics **448**, Springer, Berlin, 1975. MR 53 #1051 Zbl 0311.47021

[Miao et al. 2012] C. Miao, J. Zhang, and J. Zheng, "A note on the cone restriction conjecture", *Proc. Amer. Math. Soc.* **140**:6 (2012), 2091–2102. MR 2888196 Zbl 1250.42052

[Miao et al. 2013a] C. Miao, J. Zhang, and J. Zheng, "Linear adjoint restriction estimates for paraboloid", preprint, 2013.

[Miao et al. 2013b] C. Miao, J. Zhang, and J. Zheng, "Strichartz estimates for wave equation with inverse square potential", preprint, 2013. To appear in *Commun. Contemp. Math.*

[Planchon et al. 2003a] F. Planchon, J. G. Stalker, and A. S. Tahvildar-Zadeh, "Dispersive estimate for the wave equation with the inverse-square potential", *Discrete Contin. Dyn. Syst.* **9**:6 (2003), 1387–1400. MR 2005h:35212 Zbl 1047.35081

[Planchon et al. 2003b] F. Planchon, J. G. Stalker, and A. S. Tahvildar-Zadeh, "L^p estimates for the wave equation with the inverse-square potential", *Discrete Contin. Dyn. Syst.* **9**:2 (2003), 427–442. MR 2003j:35195 Zbl 1031.35092

[Shao 2009a] S. Shao, "A note on the cone restriction conjecture in the cylindrically symmetric case", *Proc. Amer. Math. Soc.* **137**:1 (2009), 135–143. MR 2010f:35222 Zbl 1198.42009

[Shao 2009b] S. Shao, "Sharp linear and bilinear restriction estimates for paraboloids in the cylindrically symmetric case", *Rev. Mat. Iberoam.* **25**:3 (2009), 1127–1168. MR 2011e:35362 MR 1183:42010

[Stein 1979] E. M. Stein, "Some problems in harmonic analysis", pp. 3–20 in *Harmonic analysis in Euclidean spaces, Part 1* (Williamstown, MA, 1978), edited by G. Weiss and S. Wainger, Proc. Sympos. Pure Math. **35**, Amer. Math. Soc., Providence, RI, 1979. MR 80m:42027 Zbl 0445.42006

[Stein 1993] E. M. Stein, *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, Princeton Mathematical Series **43**, Princeton University Press, Princeton, NJ, 1993. MR 95c:42002 Zbl 0821.42001

[Stein and Weiss 1971] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Mathematical Series **32**, Princeton University Press, Princeton, NJ, 1971. MR 46 #4102 Zbl 0232.42007

[Stempak 2000] K. Stempak, "A weighted uniform L^p-estimate of Bessel functions: a note on a paper of K. Guo: 'A uniform L^p estimate of Bessel functions and distributions supported on S^{n-1}'", *Proc. Amer. Math. Soc.* **128**:10 (2000), 2943–2945. MR 2000m:33004 Zbl 1031.35092

[Strichartz 1977] R. S. Strichartz, "Restrictions of Fourier transforms to quadratic surfaces and decay of solutions of wave equations", *Duke Math. J.* **44**:3 (1977), 705–714. MR 58 #23577 Zbl 0372.35001

[Taberner 1985] B. B. Taberner, "On the restriction of the Fourier transform to a conical surface", *Trans. Amer. Math. Soc.* **292**:1 (1985), 321–333. MR 86k:42023 Zbl 0578.42006

[Tao 2001] T. Tao, "Endpoint bilinear restriction theorems for the cone, and some sharp null form estimates", *Math. Z.* **238**:2 (2001), 215–268. MR 2003a:42010 Zbl 0992.42004

[Tao 2003a] T. Tao, "Recent progress on the restriction conjecture", preprint, 2003. arXiv 0311181

[Tao 2003b] T. Tao, "A sharp bilinear restriction estimate for paraboloids", *Geom. Funct. Anal.* **13**:6 (2003), 1359–1384. MR 2004m:47111 Zbl 1068.42011

[Tao et al. 1998] T. Tao, A. Vargas, and L. Vega, "A bilinear approach to the restriction and Kakeya conjectures", *J. Amer. Math. Soc.* **11**:4 (1998), 967–1000. MR 99f:42026 Zbl 0924.42008

[Vazquez and Zuazua 2000] J. L. Vazquez and E. Zuazua, "The Hardy inequality and the asymptotic behaviour of the heat equation with an inverse-square potential", *J. Funct. Anal.* **173**:1 (2000), 103–153. MR 2001j:35122 Zbl 0953.35053

[Watson 1944] G. N. Watson, *A treatise on the theory of Bessel functions*, 2nd ed., Cambridge University Press, Cambridge, 1944. MR 6,64a Zbl 0063.08184

[Wolff 2001] T. Wolff, "A sharp bilinear cone restriction estimate", *Ann. of Math.* (2) **153**:3 (2001), 661–698. MR 2002j:42019 Zbl 1125.42302

Received April 30, 2012.

JUNYONG ZHANG
DEPARTMENT OF MATHEMATICS
BEIJING INSTITUTE OF TECHNOLOGY
BEIJING 100081
CHINA
zhang_junyong@bit.edu.cn

JIQIANG ZHENG
THE GRADUATE SCHOOL OF CHINA
ACADEMY OF ENGINEERING PHYSICS
P. O. BOX 2101
BEIJING 100088
CHINA
zhengjiqiang@gmail.com

CONTENTS

Volume 264, no. 1 and no. 2

Guidelines for Authors

Authors may submit manuscripts at msp.berkeley.edu/pjm/about/journal/submissions.html and choose an editor at that time. Exceptionally, a paper may be submitted in hard copy to one of the editors; authors should keep a copy.

By submitting a manuscript you assert that it is original and is not under consideration for publication elsewhere. Instructions on manuscript preparation are provided below. For further information, visit the web address above or write to pacific@math.berkeley.edu or to Pacific Journal of Mathematics, University of California, Los Angeles, CA 90095–1555. Correspondence by email is requested for convenience and speed.

Manuscripts must be in English, French or German. A brief abstract of about 150 words or less in English must be included. The abstract should be self-contained and not make any reference to the bibliography. Also required are keywords and subject classification for the article, and, for each author, postal address, affiliation (if appropriate) and email address if available. A home-page URL is optional.

Authors are encouraged to use LaTeX, but papers in other varieties of TeX, and exceptionally in other formats, are acceptable. At submission time only a PDF file is required; follow the instructions at the web address above. Carefully preserve all relevant files, such as LaTeX sources and individual files for each figure; you will be asked to submit them upon acceptance of the paper.

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used but tags will be converted to the house format (see a current issue for examples).

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Figures prepared electronically should be submitted in Encapsulated PostScript (EPS) or in a form that can be converted to EPS, such as GnuPlot, Maple or Mathematica. Many drawing tools such as Adobe Illustrator and Aldus FreeHand can produce EPS output. Figures containing bitmaps should be generated at the highest possible resolution. If there is doubt whether a particular figure is in an acceptable format, the authors should check with production by sending an email to pacific@math.berkeley.edu.

Each figure should be captioned and numbered, so that it can float. Small figures occupying no more than three lines of vertical space can be kept in the text ("the curve looks like this:"). It is acceptable to submit a manuscript will all figures at the end, if their placement is specified in the text by means of comments such as "Place Figure 1 here". The same considerations apply to tables, which should be used sparingly.

Forced line breaks or page breaks should not be inserted in the document. There is no point in your trying to optimize line and page breaks in the original manuscript. The manuscript will be reformatted to use the journal's preferred fonts and layout.

Page proofs will be made available to authors (or to the designated corresponding author) at a website in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

PACIFIC JOURNAL OF MATHEMATICS

Volume 264 No. 2 August 2013

0030-8730(201308)264:2;1-2